LAYERED WYNER-ZIV VIDEO CODING:

A NEW APPROACH TO VIDEO COMPRESSION AND DELIVERY

A Dissertation

by

QIAN XU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Electrical Engineering

LAYERED WYNER-ZIV VIDEO CODING:

A NEW APPROACH TO VIDEO COMPRESSION AND DELIVERY

A Dissertation

by

QIAN XU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Zixiang Xiong
Committee Members,	Edward R. Dougherty
	Costas N. Georghiades
	Andrew K. Chan
	Dmitri Loguinov
Head of Department,	Costas N. Georghiades

August 2007

Major Subject: Electrical Engineering

ABSTRACT

Layered Wyner-Ziv Video Coding:
A New Approach to Video Compression and Delivery. (August 2007)
Qian Xu, B.S., University of Science & Technology of China;
M.S., Texas A&M University
Chair of Advisory Committee: Dr. Zixiang Xiong

Following recent theoretical works on successive Wyner-Ziv coding, we propose a practical *layered* Wyner-Ziv video coder using the DCT, nested scalar quantization, and irregular LDPC code based Slepian-Wolf coding (or lossless source coding with side information at the decoder). Our main novelty is to use the base layer of a standard scalable video coder (e.g., MPEG-4/H.26L FGS or H.263+) as the decoder side information and perform layered Wyner-Ziv coding for quality enhancement. Similar to FGS coding, there is no performance difference between layered and monolithic Wyner-Ziv coding when the enhancement bitstream is generated in our proposed coder. Using an H.26L coded version as the base layer, experiments indicate that Wyner-Ziv coding gives slightly worse performance than FGS coding when the channel (for both the base and enhancement layers) is noiseless. However, when the channel is noisy, extensive simulations of video transmission over wireless networks conforming to the CDMA2000 1X standard show that H.26L base layer coding plus Wyner-Ziv enhancement layer coding are more robust against channel errors than H.26L FGS coding. These results demonstrate that layered Wyner-Ziv video coding is a promising new technique for video streaming over wireless networks.

For scalable video transmission over the Internet and 3G wireless networks, we propose a system for receiver-driven layered multicast based on layered Wyner-Ziv

video coding and digital fountain coding. Digital fountain codes are near-capacity erasure codes that are ideally suited for multicast applications because of their rateless property. By combining an error-resilient Wyner-Ziv video coder and rateless fountain codes, our system allows reliable multicast of high-quality video to an arbitrary number of heterogeneous receivers without the requirement of feedback channels. Extending this work on separate source-channel coding, we consider distributed joint source-channel coding by using a single channel code for both video compression (via Slepian-Wolf coding) and packet loss protection. We choose Raptor codes - the best approximation to a digital fountain - and address in detail both encoder and decoder designs. Simulation results show that, compared to one separate design using Slepian-Wolf compression plus erasure protection and another based on FGS coding plus erasure protection, the proposed joint design provides better video quality at the same number of transmitted packets.

To My parents

ACKNOWLEDGMENTS

My five years here at Texas A&M University have been a great experience due to the many wonderful and inspirational people with whom I have had the pleasure to work. My most earnest acknowledgment must go to my advisor Prof. Zixiang Xiong. I am fortunate to have been supervised by someone with such a great enthusiasm for the research on distributed source coding. He always steers me in the right direction, and he has been instrumental in ensuring my academic, financial, and moral well-being. None of this work would have been possible without his wise guidance and constant encouragement. I would also like to thank Prof. Edward R. Dougherty for all his help and support as my second advisor for the past two years. He has taught me so much about the genomic signal processing. I thank both of my advisors for their academic and financial support throughout my stay at Texas A&M University.

I want to take this opportunity thank my other committee members, Prof. Costas N. Georghiades, Prof. Andrew K. Chan, and Prof. Dmitri Loguinov, for their constructive advice and helpful suggestions and for taking on the task of carefully reading this dissertation. My sincere thanks also go to the professors in the wireless communications group: Prof. Krishna Narayanan, Prof. Scott Miller, Prof. Erchin Serpedin, and Prof. Deepa Kundur, for their instructions and teaching, inside and outside the classroom. I am especially grateful to Prof. Vladimir Stanković, with whom I have collaborated quite a lot during my Ph.D. research. His conceptual and technical insights into my research work have been invaluable to me.

Additionally, I would like to express my sincere gratitude to my fellow graduate students in the Multimedia Lab and the Genomic Signal Processing Lab, for sharing their insightful knowledge with me. My special thanks go out to Jianping Hua, Samuel Cheng, Yang Yang, Yong Sun, Zhixin Liu and Tim Lan, who provided invaluable

support and suggestions throughout this process.

Finally, I would like to thank my family. They deserve far more credit than I can ever give them for always being there with the unconditional love, support and encouragement whenever I needed them.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

II LAYERED WYNER-ZIV VIDEO CODING 8

A. Theoretical Background 10
1. WZC 10
2. Successive WZC 12
B. Wyner-Ziv Code Design for Jointly Gaussian Sources . . . 13
C. Layered Wyner-Ziv Video Coding 15
1. Proposed Code Design 17
D. Compression Results 23
1. Successive Refinement 23
2. Layered Coding 23
a. Correlation Modeling 24
b. LDPC Code Design for SWC 24
c. Coding Performance 25
E. Error Robustness 26
1. Against Simulated Errors in the Base Layer 26
2. Against Errors from a Qualcomm Wireless Channel Simulator 27

III WYNER-ZIV VIDEO COMPRESSION AND FOUNTAIN CODES FOR RECEIVER-DRIVEN LAYERED MULTICAST . 38

A. Introduction 38
B. Fountain Codes 41
C. Wyner-Ziv Video Coding and Fountain Codes for RLM . . 42
D. Experimental Results 46

IV DISTRIBUTED JOINT SOURCE-CHANNEL CODING OF VIDEO USING RAPTOR CODES 52

A. Introduction 52
B. Theoretical Background and Related Works 56
1. SWC 56
2. WZC 57

CHAPTER Page

3. Source-channel Coding with Decoder Side Information 57
4. Related Works 58
C. Erasure Protection Coding 59
D. Separate vs. Joint Design for Distributed Source-channel Coding 61
1. Practical SWC 62
2. Transmission over Packet Erasure Channels 63
E. Distributed Joint Source-channel Coding of Video Using Raptor Codes 65
1. Encoding 67
2. Soft-decision Decoding 69
F. Experimental Results 71
1. Coding Performance with Perfect Base Layer 72
2. Coding Performance with Corrupted Base Layer . . . 74

V CONCLUSIONS 81

REFERENCES 84

VITA 93

LIST OF TABLES

TABLE Page

I The conditional entropy, LDPC code rate, and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane after NSQ of the DC (a) and the first two AC coefficients (b) and (c) after the DCT. 34

II Our computed Slepian-Wolf limits, the actual SWC rates (given by $\frac{r}{n}$) we use in the separate design, and the corresponding $\frac{r}{n}$'s of the IRA precodes in the joint design, for the DC, AC1 and AC2 of the first GOF of Foreman. The unit measure for all entries is bit. . 78

LIST OF FIGURES

FIGURE		Page
1	Lossy source coding with side information at the decoder, i.e., WZC.	11
2	Two-stage successive refinement with identical side information at the decoders.	12
3	Block diagram of the proposed layered Wyner-Ziv video coder.	17
4	NSQ throws away both the upper significant bit planes (with nesting) and the lower significant bit planes (with quantization). The number of thrown away bit planes depends on the nesting ratio N (quantization stepsize q).	18
5	NSQ with nesting ratio $N = 4$.	20
6	Bit plane based multi-stage SWC using multi-level LDPC codes for layered WZC after the DCT and NSQ.	21
7	Illustration of successive refinement in our layered Wyner-Ziv video coder, assuming ideal SWC. (a) The operational rate-PSNR function of WZC is formed as the upper concave hull of different rate-PSNR points. (b) There is almost no performance loss between monolithic WZC and layered WZC.	31
8	The estimated correlation coefficient between the DC component of the original video and that of the side information for CIF Foreman. The horizontal axis represents the rate for the H.26L base layer.	32
9	Layered WZC of Football (top), CIF Foreman (middle), and Mother_daughter (bottom), starting from different "zero-rate" points. The sum of the rates for H.26L coding and WZC is shown in the horizontal axis.	33

FIGURE Page

10 Compared to FGS coding, Wyner-Ziv video coding offers substantial improvement in decoded video quality when the base layer (or decoder side information) suffers 1% macroblock loss for Football (top), 5% macroblock loss for Foreman (middle), and 5% macroblock loss for Mother_daughter (bottom). 35

11 Comparison of the Wyner-Ziv video coder and H.26L FGS coder when both are protected with RS-based FEC codes and transmitted over a simulated CDMA2000 1X channel for Football (top), CIF Foreman (middle) and Mother_daughter (bottom). 36

12 Error robustness performance of Wyner-Ziv video coding compared with H.26L FGS for Football when both the base layer and enhancement layer bitstreams are protected with 20% RS-based FEC and transmitted over a simulated CDMA2000 1X channel with 6% PDU loss rate. The 10th decoded frame by (a) H.26L FGS and (b) Wyner-Ziv video coding in the 7th simulated transmission. 37

13 Block diagram of the proposed system. 43

14 Transmission of different GOFs. 44

15 Coding performance of the proposed scheme for: (a) the "Foreman" sequence, (b) the "Mother_daughter" sequence for two bit rates of the base layer and two packet loss rates. 50

16 Average PSNR for 100 transmissions of the first GOF of the "Foreman" video sequence vs. packet loss rate of the base layer for: (a) LC (b) HC, and three different available bandwidths. 51

17 Block diagram of our proposed video coder with Raptor codes. DCT denotes Discrete Cosine Transform, and Q stands for quantization. 66

18 (a) The graphical representation of the proposed Raptor encoder with IRA precoding. (b) The bipartite graph of our joint Raptor decoder. 77

FIGURE Page

19 Average PSNR (in dB) performance vs. bit rate (in Kb/s) between our distributed JSCC design and separate IRA and LT design for (a) the CIF Foreman and (b) the SIF Football sequences. The base layer is generated using H.26L and the packet loss ratio of the erasure channel is 0.1. The theoretical limited is $nH(X|Y)(1+\epsilon)/C$, with n being the input code length, $H(X|Y)$ the computed Slepian-Wolf limit, $\epsilon = 0.07$, and $C = 0.9$. 79

20 Performance comparisons of the joint Raptor code design, separate IRA + LT design, separate IRA + RS design, H.26L FGS + LT, and H.26L + RS for (a) the CIF Foreman and (b) the SIF Football, as a function of the packet erasure rate. All schemes are designed for packet loss ratio 0.1. 80

CHAPTER I

INTRODUCTION

Today's standard techniques for video compression are designed for "downlink" broadcast applications with one heavy encoder and multiple light decoders. Video coding standards like MPEG [1] and H.264 [2] use motion-compensated predictive DCT to achieve high compression efficiency. The encoder is the computational workhorse of the video codec while the decoder is a relatively lightweight device operating in a "slave" mode. Therefore, they are suitable for video communications (e.g. broadcast) where encoding is done only once without any power constraint and decoding performed many times.

The growing popularity of video sensor networks, video cellular phones and webcams has generated the need for low-complexity and power-efficient multimedia systems that can handle multiple video input and output streams. For example, when a natural scene is captured by spatially separated cameras and transmitted over noisy channels to a central base station for decoding, a typical new scenario of "uplink" multimedia applications arises, which has very different requirements from the traditional "downlink" scenarios. For such applications, we need a video coding system with multiple low-complexity encoders and one (or more) high-complexity decoders. In addition, the system must be robust to channel errors so that the decoder at the base station can recover the scene with high fidelity using all received bitstreams.

While standard video coding techniques (e.g., MPEG [1] and H.264 [2]) provide high compression efficiency, they fail to satisfy the requirements of the above "uplink" multimedia application. This is because the heavy computation load of DCT and

---

The journal model is *IEEE Transactions on Automatic Control.*

motion estimation is put at the encoder while the decoder is a relatively lightweight device. Typically, the complexity of a standard encoder is 5 to 10 times higher than that of the decoder. Moreover, when there are channel errors or packet losses, a decoded frame at the decoder will be different from that used at the encoder, causing the problem of error drifting that will have adverse effect on subsequent frames with severe visual degradation.

Distributed source coding (DSC) is a promising technique for "uplink" applications. DSC refers to compression of two or more correlated sources that do not communicate with each other. The main issue with DSC is to achieve the same coding efficiency as with joint (e.g., DPCM) encoding. For lossless compression of two discrete correlated sources, Slepian and Wolf [3] showed the surprising result that there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint decoding is performed. For the more general case of lossy coding with side information at the decoder, Wyner and Ziv [4] showed that it generally suffers rate loss when compared to lossy coding of the source with the side information available at both the encoder and the decoder. However, one special case of the Wyner-Ziv problem is when the source X and side information Y are zero-mean and stationary Gaussian memoryless sources and the distortion metric is MSE. The minimum bit rate needed to encode X for a given distortion when Y is available only at the decoder is equal to the rate when Y is known at both sides. In other words, there is no rate loss for this quadratic Gaussian case in Wyner-Ziv coding (WZC).

To approach the Wyner-Ziv rate-distortion (R-D) function established in [4], information-theoretic approaches were presented in [5] and several practical coding schemes for ideal jointly Gaussian sources have been proposed (see the two tutorial papers [6, 7] and references therein).

Several groups have recently explored video compression based on DSC prin-

ciples. One approach targets emerging applications (e.g., "uplink" video communications from handheld devices) that demand low encoding complexity − a scenario that is the opposite of video broadcast, for which standard coders with heavy encoding are designed. Puri and Ramchandran proposed a coder in [8] that attempts to swap the encoder-decoder complexity of standard coders. Their encoder consists of the DCT, uniform quantization and trellis coding for Slepian-Wolf compression, while their decoder performs heavy-duty motion estimation. Girod *et al.* [9] also investigated distributed video coding using a relatively low-complexity turbo code based Slepian-Wolf encoder. Whereas both coders perform better than independent intraframe (e.g., motion JPEG) coding (with the lowest encoding complexity), they suffer substantial R-D penalty when compared to standard MPEG-4 and H.264 coding (with high encoding complexity, mainly due to motion estimation). Thus there is still a large gap between what WZC or DSC theory promises (to the extent of no performance loss in certain special cases when compared to joint encoding) and what practical low-complexity distributed video coders can achieve.

Another approach is to de-emphasize low-complexity encoding while focusing on error robust Wyner-Ziv video coding. For example, Sehgal *et al.* [10] discussed how coset-based Wyner-Ziv video coding can be used to alleviate the problem of prediction mismatch in DPCM-based standard video coders. Their coder is "state-free" in the sense that the decoder does not have to maintain the same states as the encoder. Girod *et al.* [9] presented a robust video transmission system by using WZC to generate parity bits for protecting an MPEG encoded bitstream of the same video; however, since their Wyner-Ziv coder outputs parity bits (as in systematic channel coding), this scheme is better categorized as systematic source/channel coding [11] − with an MPEG systematic part plus a Wyner-Ziv parity part.

In this dissertation, we present a novel *layered* video coder based on standard

video coding and successive WZC [12, 13]. Treating a standard coded video as the base layer (or side information), a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Thus our proposed layered WZC scheme is very much like MPEG-4/H.26L FGS (Fine Granularity Scalable) coding [14, 15] in "spirit" in terms of having an embedded enhancement layer with good R-D performance. However, the key difference is that the enhancement layer is generated "blindly" without knowing the base layer in WZC. This avoids the problems (e.g., error drifting/propagation) associated with encoder-decoder mismatch in standard DPCM-based coders.

Using the H.26L coded version as the base layer, the proposed layered Wyner-Ziv video coding system over noiseless channel has roughly the same R-D performance as that of H.26L FGS [15] coding, with about 0.3 dB Peak Signal-to-Noise Ratio (PSNR) loss at high rate. In addition, we use a wireless channel simulator [16] from Qualcomm Inc. that conforms to the CDMA2000 1X standard to test the robustness of our layered Wyner-Ziv coder under wireless environments *for both the base and enhancement layers*. Extensive simulations show that layered WZC is more robust against channel errors than H.26L FGS coding, offering 0.3-1.5 dB gain in average PSNR. The advantage of layered WZC is more pronounced for high motion sequences, when error drifting with H.26L FGS becomes more severe. Our results clearly demonstrate that layered Wyner-Ziv video coding is a promising new technique for video streaming over wireless networks.

For video streaming applications that distribute data to a large number of clients, we propose a video multicast system based on layered Wyner-Ziv video coding and digital fountain coding. Layered Wyner-Ziv video coding improves robustness to packet loss compared to current scalable video coders, such as MPEG-4 FGS coder,

while generating a scalable output bitstream. To reduce the decoding time and the computational complexity (the latter being crucial for power limited wireless devices), we choose digital fountain codes [17] over RS codes for error control [18]; the latter are maximum distance separable codes with order $n \log n$ encoding time and quadratic decoding time [19]. Fountain codes are sparse-graph codes that are ideally suited for multicast applications, because they are rateless in the sense of allowing a potentially limitless stream of output symbols to be generated for a given input vector. By combining an error-resilient Wyner-Ziv video coder and rateless fountain codes, our system allows reliable multicast of high-quality video to an arbitrary number of heterogeneous receivers without the requirement of feedback channels. Our simulation results show significant performance improvements over a previously reported scheme that exploits multiple description and layered coding.

Extending the separate source-channel coding scheme proposed above, we further consider distributed source-channel coding and targets at the important application of scalable video transmission over wireless networks. The idea is to use a *single* channel code for both video compression (via Slepian-Wolf coding) and packet loss protection. First, we provide a theoretical code design framework for distributed joint source-channel coding over erasure channels and then apply it to the targeted video application. The resulting video coder is based on a cross-layer design where video compression and protection are performed jointly. We choose Raptor codes – the best approximation to a digital fountain – and address in detail both encoder and decoder designs. Using the received packets together with a correlated video available at the decoder as side information, we devise a new iterative soft-decision decoder for *joint* Raptor decoding. Simulation results show that, compared to one separate design using Slepian-Wolf compression plus erasure protection and another based on FGS coding plus erasure protection, the proposed joint design provides better video

quality at the same number of transmitted packets. Our work represents the first in capitalizing the latest in DSC and near-capacity channel coding for robust video transmission over erasure channels.

The rest of the dissertation is organized as follows: In Chapter II, we present our practical layered Wyner-Ziv video coding scheme with low-density parity-check (LDPC) code based bit plane coding for Slepian-Wolf coding (SWC). The theoretical background on WZC and Wyner-Ziv code design for jointly Gaussian sources will be provided before we put forth the layered Wyner-Ziv video coding framework. The R-D performance of our system is compared both to monolithic H.26L coding and H.26L FGS coding. After that, the simulation results of video transmission over wireless networks conforming to the CDMA2000 1X standard will be presented to show its superior error robustness over H.26L FGS coding. In Chapter III, we consider the problem of reliable multimedia delivery over the Internet and 3G wireless networks. We first give a brief overview of the digital fountain codes, followed by a step-by-step description of our proposed video multicast system by combining layered Wyner-Ziv video coding and digital fountain coding. In the end, the simulation results are presented to show significant performance improvements by our proposed system over a previously reported scheme that exploits multiple description and layered coding. A distributed source-channel coding scheme that used a single channel coding for both video compression and packet loss protection is discussed in Chapter IV. We first give theoretical background on source-channel coding with decoder side information and erasure protection coding techniques and then point out advantages of a joint source-channel code design over a separate one. The details of our proposed video coder based on Raptor codes are explained and the experimental comparisons between the proposed joint design and other separate designs are also presented. Final conclusions are drawn in Chapter V. We summarize the dissertation on the accomplished works

and provide a perspective for the future research in distributed video coding.

CHAPTER II

LAYERED WYNER-ZIV VIDEO CODING *

In this chapter, we present a novel *layered* video coder [20] based on standard video coding and successive WZC [12, 13]. Treating a standard coded video as the base layer and decoder side information, our layered Wyner-Ziv encoder consists of DCT, nested scalar quantization (NSQ), and irregular LDPC code based SWC [21]. The DCT is applied as an approximation to the conditional Karhunen-Loeve transform (cKLT) [22], which makes the components of the transformed block conditionally independent given the decoder side information. NSQ is a binning scheme that facilitates layered bit plane coding of the bin indices while reducing the bit rate. SWC plays the role of conditional entropy coding (with side information at the decoder) for further compression. Our Wyner-Ziv decoder performs joint decoding by combining the base layer and the Wyner-Ziv bitstream for enhanced video quality

We aim to generate a layered Wyner-Ziv bitstream from the *original video sequence* such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Thus our proposed layered WZC scheme is very much like MPEG-4/H.26L FGS [14, 15] and H.263+ coding [23] in "spirit" in terms of having an embedded enhancement layer with good R-D performance. However, the key difference is that the enhancement layer is generated "blindly" without using the base layer in WZC. This alleviates the problem of error drifting/propagation associated with encoder-decoder mismatch in standard DPCM-based coders because, as we shall see later in Section E, a corrupted base layer (with errors within a certain range) can

*©[2006] IEEE. Reprinted, with permission, from "Layered Wyner-Ziv video coding" by Q. Xu and Z. Xiong, 2006. *IEEE Transactions on Image Processing*, vol. 15, pp. 3791-3803.

still be combined with the enhancement layer for Wyner-Ziv decoding. This inherent error-resilience in the base layer is the main advantage of our proposed Wyner-Ziv coder over standard FGS coding.

Our layered WZC scheme has the attractive feature that encoding is done only once but decoding allowed at many lower bit rates with commensurate qualities, i.e., there is no performance difference between layered and monolithic WZC for the enhancement layer. This is because our work is underpinned by recent theoretical results [12, 13] that extend the successive refinability of Gaussian sources from classic source coding to WZC and because our design is based on scalar quantization and bit plane coding. While the code design in [13] assumes ideal Gaussian sources with MSE distortion, results here are the first reported on practical layered WZC of video that *do not* suffer performance loss due to layering in WZC. The conference version of this chapter appeared in [24]. Other groups' works on scalable Wyner-Ziv video coding are [25, 26, 27].

Relying on the H.26L coded version as the base layer[1], our Wyner-Ziv video coder is capable of achieving roughly the same R-D performance as the H.26L FGS coder in [15]. For example, using irregular LDPC codes of lengths in the order of 8×10^4 bits for SWC, the former performs 0.3 dB worse in average PSNR than the latter at high rate. The performance of our layered Wyner-Ziv coder only degrades slightly when the LDPC code lengths are decreased by a factor of four to reduce latency.

Besides the thrust of applying the theory of successive WZC to practical video compression and showing the competitiveness of layered Wyner-Ziv video coding with standard FGS coding, this chapter also aims to highlight error robustness of the base layer due to Wyner-Ziv enhancement layer coding via simulations, where a fixed

[1]H.26L refers to the video codec of the now well-known H.264 standard [2], which is the result of the combined efforts of ITU and ISO MPEG.

amount of macroblock loss is introduced to the H.26L coded base layer. After combining the corrupted base layer with enhancement layers generated from WZC instead of FGS coding, we observe an improvement of 0.6-2.71 dB in video quality measured in average PSNR. This means that in video streaming applications, error-free delivery of the base layer is less critical with our layered Wyner-Ziv video coder than with standard scalable coders (e.g., MPEG-4/H.26L FGS).

Finally, we use a wireless channel simulator [16] from Qualcomm Inc. that conforms to the CDMA2000 1X standard to test the robustness of our layered Wyner-Ziv coder under wireless environments *for both the base and enhancement layers*. Extensive simulations show that layered WZC is more robust against channel errors than H.26L FGS coding, offering 0.3-1.5 dB gain in average PSNR. In addition, the advantage of layered WZC is more pronounced for high motion sequences, when error drifting with H.26L FGS becomes more severe. Our results clearly demonstrate that layered Wyner-Ziv video coding is a promising new technique for video streaming over wireless networks.

The rest of the chapter is organized as follows. Section A gives the theoretical background on WZC. Section B covers Wyner-Ziv code design for jointly Gaussian sources. Section C puts forth our layered Wyner-Ziv video coding framework. Section D presents our video compression results, while Section E focuses on error robustness of our layered Wyner-Ziv video coder.

A. Theoretical Background

1. WZC

Consider $\{(X_i, Y_i)\}_{i=1}^{\infty}$ as a sequence of independent drawings of a pair of correlated discrete random variables X and Y. The problem of separate lossless encoding and

joint decoding of X and Y was first considered by Slepian and Wolf [3], who gave the achievable rate region as

$$R_X \geq H(X|Y),\ \ R_Y \geq H(Y|X),\ \ R_X + R_Y \geq H(X,Y).$$

This surprising result indicates that the joint entropy $H(X,Y)$ is still achievable. Hence there is no performance loss with SWC when compared to joint encoding. But the caveat is that SWC is only lossless asymptotically with respect to the code length.

Lossless source coding with side information at the decoder is a special case of the SWC problem. Assume Y is encoded with $H(Y)$ bits so that it can be perfectly decoded at the decoder, then according to (1), SWC of X boils down to compressing it to the rate limit $H(X|Y)$. We henceforth limit ourselves to equating SWC with lossless source coding with side information in this chapter.

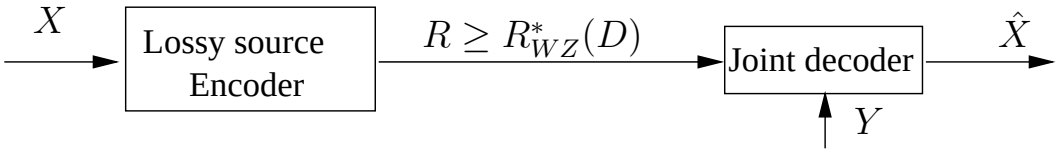


Fig. 1. Lossy source coding with side information at the decoder, i.e., WZC.

WZC [4] generalizes the setup of SWC in that coding of X is with respect to a fidelity criterion rather than lossless (as depicted in Fig. 1). In addition, the source X could be either discrete or continuous. The work of [4] examines the question of how many bits are needed to encode the source X under the constraint that the average distortion between X and decoded version $\hat{X}$ satisfies $E\{d(X,\hat{X})\} \leq D$, assuming that the side information Y (discrete or continuous) is available only at the decoder. Denote $R^*_{WZ}(D)$ as the achievable lower bound of the bit rate for an expected distortion D for WZC, and $R^*_{X|Y}(D)$ as the R-D function of coding X with

side information Y available also at the encoder.

In general there is a rate loss associated with WZC, that is: $R^*_{WZ}(D) \geq R^*_{X|Y}(D)$. However, $R^*_{WZ}(D) = R^*_{X|Y}(D)$ when X and Y are zero-mean and jointly Gaussian and the distortion measure is MSE [4] [2]. We restrict ourselves to this jointly Gaussian case in WZC because there is no rate loss and it is of special interest in practice, where many image and video sources can be modeled as jointly Gaussian after mean subtraction.

2. Successive WZC

A successive refinement code for the Wyner-Ziv problem consists of multi-stage encoders and decoders where each decoder uses all the information generated from decoders of its earlier stages [12]. Fig. 2 depicts a special case of two-stage successive coding for the Wyner-Ziv problem with the side information at each stage being the same.

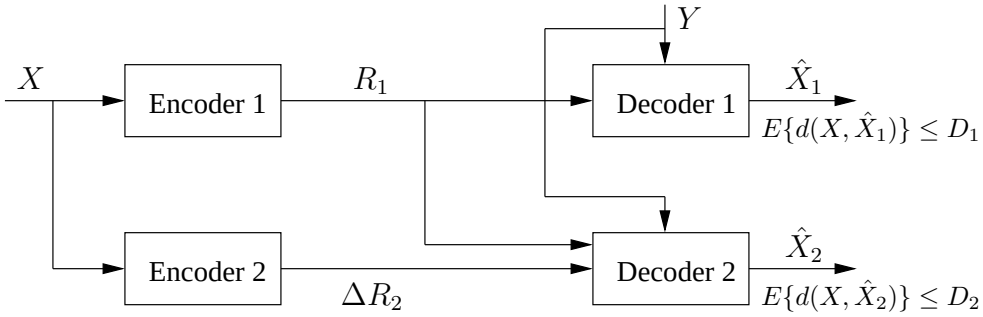


Fig. 2. Two-stage successive refinement with identical side information at the decoders.

Let Y be the side information available to the decoder at both the coarse stage and the refinement stage, and the corresponding coding rates (distortions) are $R_1(D_1)$

[2]Recently Pradhan *et al.* [28] extended this no rate loss result to the more general case with $X = Y + Z$, where Y and Z are independent and only Z is i.i.d. Gaussian (Y can follow arbitrary distribution).

and $R_2(D_2)$, respectively. A source X is said to be *successively refinable* from D_1 to D_2 ($D_1 > D_2$) with side information Y if

$$R_1 = R^*_{WZ}(D_1) \quad \text{and} \quad R_1 + \Delta R_1 = R^*_{WZ}(D_2). \tag{2.1}$$

The notion of successive coding can be naturally extended to any finite number of stages [12]. Consider the case when the side information fed into the K decoders at each level is the same, the source X is *multi-stage successively refinable* with side information Y if

$$R_1 = R^*_{X|Y}(D_1) \quad \text{and} \quad R_i + \Delta R_i = R^*_{X|Y}(D_{i+1}), \tag{2.2}$$

for $i = 1, 2, \ldots k-1$.

Necessary and sufficient conditions for successive refinability are given in [12] and the jointly Gaussian source (with MSE measure) shown to be multi-stage successively refinable in the Wyner-Ziv setting. Extending the successive refinement result of [12] on jointly Gaussian sources, Cheng and Xiong [13] proved that the jointly Gaussian condition can be relaxed to the case that only the difference $X-Y$ between the source X and the side information Y is Gaussian and independent of the side information Y, i.e., the more general class of sources without rate loss in WZC defined in [28] is also successively refinable.

B. Wyner-Ziv Code Design for Jointly Gaussian Sources

Wyner-Ziv code design involves both source coding (quantization) and channel coding [7]. The simplest Wyner-Ziv coder involves a 1-D nested lattice/uniform quantizer [5] with a coarse coset code nested in a fine coset code, i.e., the coarse code is a subcode of the fine code. The fine code does source coding while each coarse coset

code performs channel coding. This coset coding scheme amounts to binning, which refers to dividing the space of all possible outcomes of a source into disjoint subsets (or bins). To encode, X is first quantized by the fine source code, resulting in quantization errors, then only the index of the bin that the quantized X belongs to is coded to save the rate. Using this coded index, the decoder finds in the bin (or coset code) the codeword closest to the side information Y as the best estimate of X. There is quantization error due to source coding and binning loss due to channel coding.

Usually, there is still correlation remaining in the quantized version of X and the side information Y, and SWC can be employed to exploit this correlation to reduce rate. Thus SWC is an integral part of WZC, much like the way entropy coding is used in classic lossy source coding to achieve further compression after quantization.

The connection between SWC and channel coding was first made in [29] and an explicit syndrome-based binning scheme was outlined there based on parity-check codes. The idea is to partition the 2^n possible source inputs (assuming binary source X with block coding of length n) into 2^{n-k} bins, each with 2^k elements, and index them with syndromes of a binary (n, k) parity-check code. The bin with zero syndrome corresponds to all valid codewords of the channel code, and the rest are different shifted versions of it. This way the distance property of the channel code is preserved among elements in each bin. Encoding only involves multiplying the $(n-k)\times n$ parity check matrix of the channel code with the n-bit input sequence and outputting the $n-k$ syndrome bits that index the bin to which the input sequence belongs. The resulting compression ratio is $n : n-k$. The decoder takes the bin index and finds the element closest to the side information in the bin as the best estimate of the input sequence. This syndrome-based binning scheme can approach the rate limit $H(X|Y) = 1 - r$ of SWC if a near-capacity parity-check code with rate $r = \frac{k}{n}$ is designed for the channel that characterizes the correlation between X and Y. The

first practical design that follows this scheme using LDPC codes [30] was reported in [21], showing performance very close to the Slepian-Wolf limit $H(X|Y)$.

For WZC of jointly Gaussian sources, the limit-approaching Slepian-Wolf code design in [21] can be combined with strong source codes (e.g., TCQ [31]) to approach the theoretical limit. This forms the base of the Slepian-Wolf coded quantization paradigm [7] for WZC that generalizes entropy-coded quantization for classic source coding. The role of SWC is to approach the rate limit $H(Q(X)|Y)$, where $Q(X)$ is the quantized version of the input Gaussian source X. This requires a simple extension of the syndrome-based binning scheme [29] from SWC of binary sources to M-ary sources with multi-level LDPC codes.

According to [7], the performance gap of high-rate Slepian-Wolf coded quantization to the Wyner-Ziv distortion-rate (D-R) function is exactly the same as that of high-rate classic source coding to the D-R function. A practical layered Wyner-Ziv code design based on NSQ and multi-level LDPC codes for SWC was presented in [13], yielding results that are 2.9 to 1.65 dB away from the Wyner-Ziv D-R function for rates ranging from 0.48 to 6.0 b/s.

C. Layered Wyner-Ziv Video Coding

Successive or scalable image/video coding made popular by EZW [32] and 3-D SPIHT [33] is attractive in practical applications such as networked multimedia. By producing a video stream that can be decoded at more than one quality levels, scalable video coding achieves graceful quality degradation as the available bandwidth for data transmission decreases. This is very desirable in video streaming applications.

When decoder side information is available, it is also important and rewarding to explore successive Wyner-Ziv video coding in practice. Although practical code

designs in [7, 13] for WZC of jointly Gaussian sources perform close to the theoretical limit, it is not straightforward to apply these designs directly to video sources. The first issue in Wyner-Ziv video coding is the identification of the decoder side information. Our novelty is to use a standard decoded low-quality video as the side information, which is highly correlated with the original video source. In addition, there are several other issues involved in Wyner-Ziv video coding.

Transform design: Unlike i.i.d. Gaussian sources, the neighboring pixels in a video frame are highly correlated with each other. In standard video coding, the DCT has been widely used to decorrelate the image pixels to facilitate compression. For Wyner-Ziv video coding, ideally the cKLT [22] should be applied to both the video source and the side information to make the former conditionally independent given the latter before performing WZC. But the cKLT is signal-dependent, in practice a signal-independent approximation has to be used.

Correlation modeling: In WZC of jointly Gaussian sources, the joint statistics of the sources is assumed to be known *a priori*. In Wyner-Ziv video coding, the source correlation depends on the video quality of the side information. In practice it has to be estimated via correlation modeling, which is a critical step as it directly determines the performance limit of WZC.

Quantization: To approach the Wyner-Ziv limit, strong quantizers such as TCQ have to be employed in conjunction with limit-approaching Slepian-Wolf codes. However, TCQ does not facilitate successive refinement (although it can lead to progressive coding). Thus we are confined to NSQ that allows bit plane coding for successive refinement. Fortunately, the performance loss of using NSQ for WZC of jointly Gaussian sources is only 1.53 dB at high rate (assuming ideal SWC) [7].

Slepian-Wolf code design and rate control: Capacity-achieving channel codes such as LDPC codes [30] have to be used to approach the Slepian-Wolf limit. However,

to achieve high performance with these advanced channel codes requires long block length, which is not a problem with jointly Gaussian sources but introduces long delay in video coding. In addition, the code rate for SWC and convergence at the Slepian-Wolf decoder heavily rely on the correlation between the source and the side information. Our main contribution here lies in the design of efficient multi-level LDPC codes to realize layered Wyner-Ziv video coding via SWC of successive bit planes after NSQ, starting from the most significant bit plane.

1. Proposed Code Design

We now present our layered Wyner-Ziv video coder using LDPC code based SWC. Treating a standard H.26L decoded video as the base layer (and side information), a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Denote the current frame of the original video as $\mathbf{x}$, which is encoded with H.26L to obtain the base layer (or side information) $\mathbf{y}$. Fig. 3 depicts the block diagram of our layered Wyner-Ziv coder, whose encoder consists of three components: the DCT, NSQ, and SWC based on irregular LDPC codes.

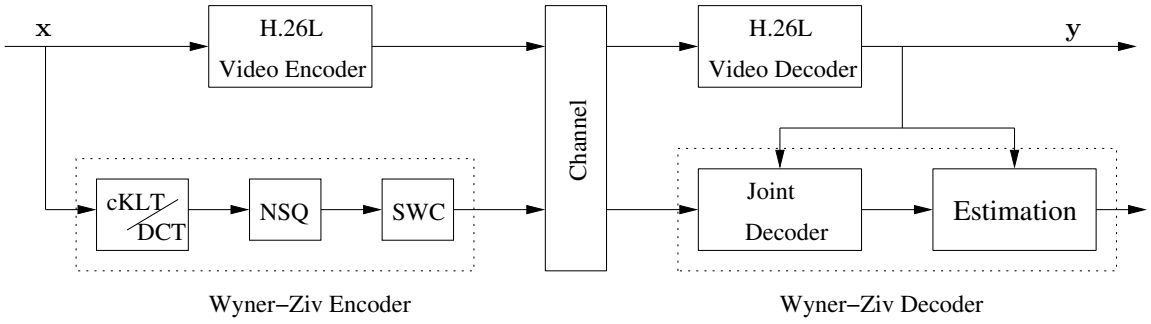


Fig. 3. Block diagram of the proposed layered Wyner-Ziv video coder.

We use the DCT as an approximation to the cKLT [22], which makes the coeffi-

cients of the transformed block of the original video $\mathbf{x}$ conditionally independent given the same transformed block of the side information $\mathbf{y}$. NSQ is a binning scheme that assigns the input DCT coefficients $\mathbf{X}$ to cosets and outputs only the coset indices. The DCT coefficients are split into several bit planes with their binary representations. The upper significant bit planes of the DCT coefficients are skipped in NSQ since they are highly correlated to those in the side information. There will be a significant loss in quality if the side information cannot be used to correctly recover these bit planes at the joint Wyner-Ziv decoder. The lower significant bit planes are less important and hence quantized to zero by NSQ to save rate. Therefore, both the upper and lower significant bit planes are thrown away in NSQ depending on the nesting ratio N and quantization stepsize q, respectively, and only those in between are coded (see Fig. 4).

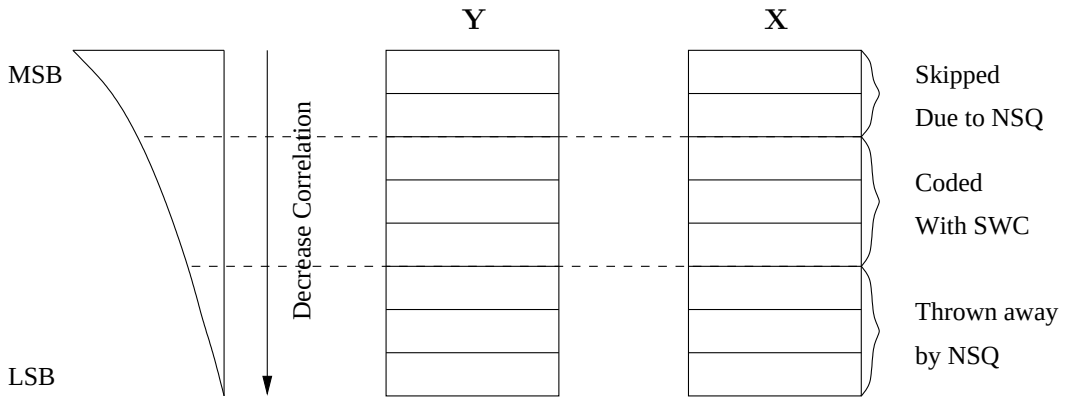


Fig. 4. NSQ throws away both the upper significant bit planes (with nesting) and the lower significant bit planes (with quantization). The number of thrown away bit planes depends on the nesting ratio N (quantization stepsize q).

NSQ introduces both a binning loss, which should be kept small with strong channel coding, and a quantization loss that should be optimally traded off with rate in source coding. In addition, there is still correlation between the quantized version (bit planes in the middle) of the source $\mathbf{X}$ and the side information $\mathbf{Y}$, and SWC can

be employed to exploit this correlation and achieve further compression. We employ multi-level LDPC codes for SWC in the third component of the encoder and output one layer of compressed bitstream for each bit plane after NSQ. In doing so, we note that the correlation decreases as we move from the most significant bit (MSB) to the least significant bit (LSB). Thus higher rate LDPC codes are designed for higher bit planes to achieve more compression; while lower rate LDPC codes are given to lower bit planes for less compression. Furthermore, although theoretically the overall rate required is the same with different orders of bit plane coding, to facilitate layered coding, the order of encoding proceeds from the MSB to the LSB after NSQ. In the following, we will explain each component in details.

Transform: For WZC of $\mathbf{x}$, we first apply the cKLT (approximated by the DCT) to every 4×4 block of $\mathbf{x}$ so that the components of the transformed block $\mathbf{X} = \mathbf{T}\mathbf{x}$ are conditionally independent given the side information $\mathbf{y}$, which is also transformed into $\mathbf{Y} = \mathbf{T}\mathbf{y}$. Each frequency component of $\mathbf{Y}$ (denoted by Y) acts as the side information for the corresponding component of $\mathbf{X}$ (denoted by X). We assume that X and Y are jointly Gaussian with $X = Y + Z$, where Z is zero-mean Gaussian and independent of Y (although DCT coefficients of images/video are better modeled as Laplacian distributed [34]).

Quantization: The next step is NSQ (see Fig. 5), which consists of a coarse coset channel code with minimum distance $d_{min} = Nq$ nested in a fine uniform scalar quantizer with stepsize q. To encode, X is first quantized by the fine source code (uniform quantizer), resulting the "good" distortion, which is the average quantization error of $q^2/12$ at high rate. However, only the index B ($0 \leq B \leq N-1$) of the coset in the coarse channel code that the quantized X belongs to is coded to save rate. Using the decoded coset index B, the decoder finds in the coset the codeword closest to the side information Y as the best estimate of X. Due to the coset channel code

employed in nesting process, the Wyner-Ziv decoder suffers a small probability of "bad" distortion that is inversely proportional to $d_{min} = Nq$. It is desirable to choose a small quantization stepsize q to minimize the "good" distortion due to source coding. On the other hand, d_{min} should be maximized to minimize the "bad" distortion due to channel coding. Thus for a fixed N, there exists an optimal q that minimizes the total distortion, which is the sum of the "good" distortion and the "bad" distortion.

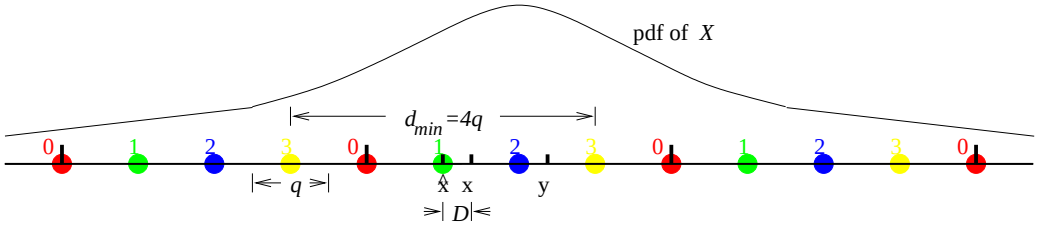


Fig. 5. NSQ with nesting ratio $N = 4$.

SWC: Due to the correlation between X and Y, there still remains correlation between the coset index B and the side information Y. SWC can be used to compress B to the rate of $R = H(B|Y)$. Express B in its binary representation as $B = B_0 B_1 \dots B_{m-1}$, where B_0 is the MSB, B_{m-1} is the LSB, and $m = \lceil \log_2 N \rceil$. We employ multi-level LDPC codes to compress $B_0 B_1 \dots B_{m-1}$ based on the syndrome-based approach [21, 29]. The rate of the LDPC code for B_i $(0 \le i \le m-1)$ depends on the conditional entropy $H(B_i|B_0, \dots, B_{i-1}, Y)$ [13], which denotes the minimum rate needed for lossless recovery of B_i given $B_0 \dots B_{i-1}$ and Y at the decoder.

We initially assume ideal SWC in the sense that the rate $R = H(B|Y)$ can be achieved. Then for each fixed N (number of cosets in the channel code), we vary the uniform quantization step size q to generate a set of R-D points (R, D) and pick the optimal q^* corresponding to the point with the steepest R-D slope from the zero-rate point in WZC. Note that the distortion for the zero-rate point is just $||X - Y||^2$, which is the average distortion of base layer coding due to H.26L. After

identifying the optimal R-D points for different N, the lower convex hull of these points forms the operational R-D curve of WZC. Due to the fact that quadratic Gaussian sources are successively refinable [12, 13], the same operational R-D curve should be traversed[3] by starting with a large N (with its corresponding q^*) first and then sequentially dropping bit planes of B. In other words, by setting different low bit plane levels of B to zero, the resulting R-D points after Wyner-Ziv decoding should all lie on the operational R-D curve. Our simulations verify this property of successive refinement and justify our approach of coding B_i into the i-th layer with rate $H(B_i|B_0, \ldots, B_{i-1}, Y)$ (see Fig. 6). By the chain rule $H(B|Y) = H(B_0|Y) + H(B_1|B_0, Y) + \ldots + H(B_{m-1}|B_0, \ldots B_{m-2}, Y)$. So layered coding suffers no rate loss when compared with monolithic coding.

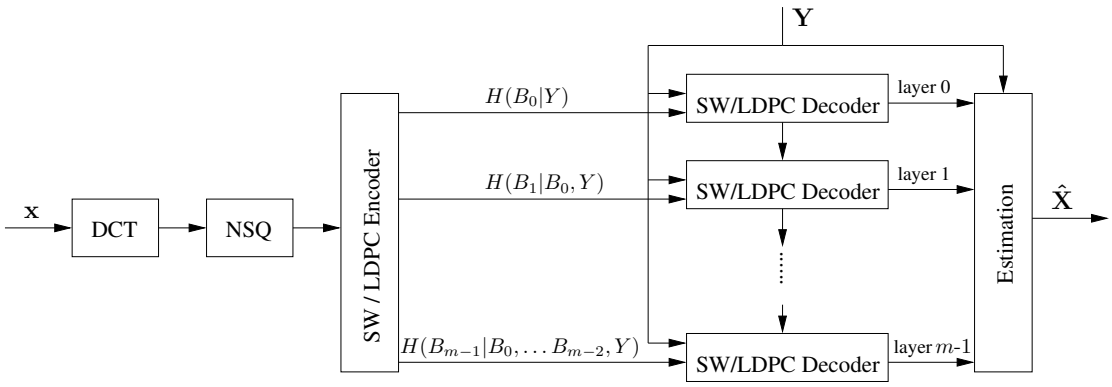


Fig. 6. Bit plane based multi-stage SWC using multi-level LDPC codes for layered WZC after the DCT and NSQ.

In our irregular LDPC code designs, the code degree distribution polynomials $\lambda(x)$ and $\rho(x)$ of the LDPC codes are optimized using density evolution based on the Gaussian approximation [35]. The bipartite graph for the irregular LDPC code is then randomly constructed based on the optimized code degree distribution polynomials

[3]Here we assume that X and Y are jointly Gaussian, even though this is only approximately true.

$\lambda(x)$ and $\rho(x)$. To compress bit plane B_i, only the corresponding syndrome determined by the sparse parity check matrix of the irregular LDPC code is transmitted to the decoder. At the decoder, each additional bitstream/syndrome layer is combined with previously decoded bit planes to decode a new bit plane before joint estimation of the output video. Let $\hat{B}_i$ represent the reconstruction of B_i. The message-passing algorithm [36] is used for iterative LDPC decoding, in which the received syndrome bits correspond to the check nodes on the bipartite graph, the side information and the previously decoded bit planes provide the *a priori* information as to how much is the probability that the current bit is "1" or "0", i.e., $LLR = \log \frac{p(B_i=0|\hat{B}_0,\ldots,\hat{B}_{i-1},Y)}{p(B_i=1|\hat{B}_0,\ldots,\hat{B}_{i-1},Y)}$.

After decoding B_0 as $\hat{B}_0$, both $\hat{B}_0$ and Y will be fed into the decoder for decoding of B_1. Since the allocated bit rate for coding B_1 is $H(B_1|B_0, Y)$, B_1 can be correctly decoded as long as $\hat{B}_0 = B_0$. By multi-stage decoding, B_i can be correctly recovered with the help of Y and the previously decoded bit planes $B_0, B_1, \cdots, B_{i-1}$, which are already available at the decoder. The more syndrome layers the decoder receives (or the higher the bit rate), the more bit planes of B will be recovered to better reconstruct X. Therefore, successive WZC provides the flexibility to accommodate a wide range of bit rates. Progressive decoding is desirable for applications where only a coarse description of the source suffices at the first stage with low bit rate, and fine details are needed at some later stage with higher bit rate.

We perform optimal estimation at the joint decoder. The decoded coset index $\hat{B}_0\hat{B}_1 \ldots \hat{B}_i$ specifies the uncertainty region of X. The side information essentially supplies the conditional pdf of X given Y, which is a Gaussian with mean Y and variance proportional to the correlation between Y and X. The optimal estimate of X is computed as the conditional centroid $\hat{X} = E(X|\hat{B}_0\hat{B}_1 \ldots \hat{B}_i, Y)$. Finally, the inverse DCT is applied to $\hat{\mathbf{X}}$ to obtain $\hat{\mathbf{x}}$ in the pixel domain.

D. Compression Results

1. Successive Refinement

Due to the approximation of the cKLT by the DCT and the Gaussian assumption of X and Y in our practical Wyner-Ziv video coder, experiments are carried out on the CIF Foreman sequence to verify the validity of our practice and illustrate successive refinement by assuming ideal SWC, that is the rate $R = H(B|Y)$.

The input video is first encoded with H.26L to obtain base layer (and the side information). Then the proposed Wyner-Ziv video coding scheme which consists of the DCT, NSQ and *ideal* bit plane based SWC (with rate $R = H(B|Y)$) is used to generate a Wyner-Ziv bitstream to enhance the base layer. The rate-PSNR performance for four different values of $N \in \{2, 4, 8, 16\}$ with different q's for each N, starting from two different zero-rate points on the base layer rate-PSNR performance, is plotted in Fig. 7 (a). After that, the operational rate-PSNR function of WZC is formed as the upper concave hull of different rate-PSNR points. Then starting at a high rate point on the operational rate-PSNR function in Fig. 7 (a) (e.g. with $N = 16$ and its corresponding q^*), we perform layered coding by dropping more and more lower bit planes of the coset index B to achieve lower rates. Fig. 7 (b) shows good match between the performance of monolithic WZC and that of layered WZC.

2. Layered Coding

We implement SWC based on irregular LDPC codes and investigate the layered WZC performance for Football (352×240) and CIF Foreman and Mother_daughter, since these sequences represent different amount of motion. Standard H.26L encoded video is treated as side information at the decoder. One hundred frames are compressed with a frame rate of 30 Hz. For each of these sequences, the first frame is coded as

I frame, and all the subsequent frames as P frames by H.26L. Different quantization stepsizes are used in the H.26L coder to generate different "zero-rate" points for WZC. After DCT of the original video, different transform coefficients are encoded independently, and we only code the first three DCT coefficients.

a. Correlation Modeling

We assume that $X = Y + Z$ in the DCT domain, where the side information $Y \sim N(0, \sigma_Y^2)$ and the quantization noise $Z \sim N(0, \sigma_Z^2)$ due to H.26L coding are independent. We estimate σ_Z^2 separately for different DCT coefficients by computing the MSE between X and Y. Fig. 8 shows our empirically estimated correlation coefficient $\sqrt{1 - \sigma_Z^2/\sigma_X^2}$ between the DC component of the original video X and that of the side information Y for Foreman.

b. LDPC Code Design for SWC

Recall that we only apply NSQ to the first three DCT coefficients and compress them with SWC. For each transform coefficient, we use a four-bit nested scalar quantizer to generate four bit planes. The 12 bit planes are then encoded by 12 different LDPC codes (of different rates). The LDPC code rate for the i-th bit plane B_i is maximized to approach the conditional entropy $H(B_i|B_0, \ldots, B_{i-1}, Y)$, meaning LDPC code design has to be tailored to the specific sequence or group of frames (GOF). We note that this code rate only depends on the joint statistics between B and Y (or more specifically between B_i and $\{B_0, \ldots, B_{i-1}, Y\}$) − the reason why we can encode X without using the actual value of Y.

The degree distribution polynomials of the LDPC codes are optimized using the Gaussian approximation [35] and the bipartite graphs of them are generated randomly. As an example, for CIF Foreman with 530 Kbps for the base layer, the

optimal quantization stepsize for the NSQ is $q = 32.0$ with nesting ratio $N = 16$. For the first GOF, the conditional entropy (or rate limit of SWC), LDPC code rate[4] and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane after the NSQ of the DC component and the first two AC components of the DCT coefficients are listed in Table I, which indicates more loss at lower bit planes due to practical LDPC coding when the conditional entropy is higher.

c. Coding Performance

Starting with the largest $N = 16$ and its corresponding optimal q^*, we quantize X into B and sequentially decode B_0, B_1, B_2 and B_3. When the LDPC code lengths are in the order of 8×10^4 bits, we group 20 frames together in WZC. One hundred iterations are used for LDPC iterative decoding to achieve the bit error probability of 5×10^{-5}. The same pseudo-random seed is used at both the encoder and the decoder such that the same codebooks are used. The joint decoder performs optimal estimation based on the side information Y and the decoded coset index. Compared to ideal SWC with $R = H(B|Y)$, the loss due to practical LDPC coding is 0.05 b/s. Layered WZC results in terms of rate-PSNR performance is shown in Fig. 9. We see that as more bit planes are decoded, the video quality improves. We also observe that the performance loss due to WZC rather than FGS coding decreases as the bit rate for H.26L base layer (or "zero-rate" for WZC) increases, with a maximum PSNR loss of 0.3 dB at the same rate. This is partially because the correlation between X and Y is higher when the base layer is coded at higher rate with better quality (see also Table I).

To cut the latency introduced by SWC, we reduce the LDPC code lengths to the

[4]The actual compression rate is one minus the LDPC code rate.

order of 2×10^4 bits (while using the same degree profiles as before). This way only five frames are grouped together in WZC. The rate loss in SWC is increased slightly from 0.05 to 0.09 b/s, which is translated into a maximum PSNR loss of 0.5 dB. As seen from Fig. 9, the PSNR performance loss for Foreman is small when the code length for SWC is scaled down from 8×10^4 to 2×10^4 bits.

E. Error Robustness

1. Against Simulated Errors in the Base Layer

Our layered Wyner-Ziv video coding framework is very similar to FGS coding [14, 15] in the sense that both schemes treat the standard coded video as the base layer and generate an embedded bitstream as the enhancement layer. However, the key difference is that instead of coding the difference between the original video and the base layer reconstruction as with FGS, the enhancement layer is generated "blindly" without knowing the base layer in Wyner-Ziv video coding. Therefore, the stringent requirement of FGS coding that the base layer is always available losslessly at the decoder/receiver can be loosened somewhat because an error-concealed version of the base layer can still be used in the joint Wyner-Ziv decoder. That the latter statement is true can be easily seen from the NSQ depicted in Fig. 5, since any side information $y \in (\hat{x} - d_{min}/2, \hat{x} + d_{min}/2)$ will result in the same decoded $\hat{x}$.

In our experiments, the same video sequence is compressed by both our Wyner-Ziv video coder and the H.26L FGS coder [15] at a frame rate of 30 Hz. The bit rate for the base layer is the same, so is for the enhancement layers. Every 15 frames start with one I frame, followed by 14 P frames. We introduce the same amount of macroblock loss to the base layer for both coders and compare their error robustness.

For the Football sequence, the base layer is encoded at 1450 Kb/s and the bit

rate for the enhancement layer of both the WZC and the FGS coding is 200 Kb/s (for the top two bit planes). Then 1% macroblock loss in the base layer is simulated with simple error concealment [37] performed during decoding of the base layer. The PSNRs of the first 15 frames are shown in Fig. 10. The performance of Wyner-Ziv video coding is 2.71 dB better on average than H.26L FGS coding. This is because the basic assumption of error-free delivery of the base layer in FGS coding is no longer valid in this setup while the error-concealed version of the base layer can still be used as side information in Wyner-Ziv decoding.

Results from similar experiments with CIF Foreman and Mother_daughter are also given in Fig. 10. For Foreman, the base layer is encoded at 190 Kb/s and the bit rate for the enhancement layer of both the WZC and the FGS coding is 60 Kb/s (for the MSB). 5% macroblock loss is introduced to the base layer. The performance of Wyner-Ziv video coding is 1.93 dB better on average than H.26L FGS coding. For Mother_daughter, the base layer is encoded at 146 Kb/s and the bit rate for the enhancement layer of both the WZC and the FGS coding is 108 Kb/s (for the top two bit planes). 5% macroblock loss is introduced to the base layer. The performance of Wyner-Ziv video coding is 0.6 dB better on average than H.26L FGS coding.

2. Against Errors from a Qualcomm Wireless Channel Simulator

To test error robustness of the *base and enhancement layer* bitstream of our Wyner-Ziv coder, a wireless channel simulator [16] is obtained from Qualcomm Inc. This simulator adds packet errors to streams of real-time transport protocol (RTP) packets transmitted over wireless networks conforming to the CDMA2000 1X standard. It assumes the use of a dedicated radio channel for the RTP packet stream under a given maximum transmission rate. Furthermore, protocol data unit (PDU) losses are introduced in the radio link control (RLC) layer. Each RTP packet is fragmented into

equal-size PDUs and it is considered successfully received by the decoder only when all its PDUs are received and the arriving time of its last PDU is still within the maximum end-to-end delay. The Qualcomm simulator also provides FEC emulation with Reed-Solomon (RS) code, which is assumed to have ideal error correction capability. In RS coding, source symbols are encoded into RS parity symbols to provide protection for PDU losses.

In our experiments, each video sequence is compressed into slice RTP packets by both our Wyner-Ziv video coder and the H.26L FGS coder [15] at a frame rate of 30 Hz. The GOF size is now set to 20 with the structure of IP$\cdots$P. The bit rate for the base layer is the same, so is for the enhancement layers. Both the base and enhancement layers are protected with RS parity packets, whose rate is specified by the overhead percentage as an input parameter to the channel simulator. We set the FEC overhead percentage to 25% so that 20% of the overall bit rate is used for RS-based FEC. The resulting RTP packet streams are transmitted over CDMA2000 1X wireless networks simulated with the Qualcomm simulator. PDU losses are introduced to the RLC layer for both coders for error robustness comparisons.

For Football, the H.26L base layer is encoded at 1470 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 400 Kb/s. Thus the total transmission rate is (1470+400)$\times$(1+25%)=2337 Kb/s. PDU loss rates of up to 10% are simulated across the *entire* packet stream on the RLC layer.

Despite 20% FEC, owing to the stringent latency constraint and packet fragmentation during transmission, there are still *residual* RTP packet losses at the decoder. For example, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are 0.15%, 0.75%, 1.76%, 3.30%, and 5.19%, respectively. Simple error concealment is performed during decoding of the base layer. As for the enhancement layer bitstream, the *whole* layer with the first packet error, together with all

subsequent layers, are discarded in the Wyner-Ziv video decoder since Slepian-Wolf decoding cannot proceed with corrupted syndromes in any layer. On the other hand, decoding the embedded FGS bitstream only stops at the first detected lost packet for the H.26L FGS coder.

For each PDU loss rate, 200 video transmissions are simulated, and Fig. 11 depicts the average PSNR performance vs. PDU loss rate for the two coders. When the base layer is perfectly reconstructed (e.g., no error occurs in the base layer in 21% of our simulated transmissions when the PDU loss rate is 6%), the H.26L FGS coder performs slightly better than the Wyner-Ziv video coder (as seen in the coding performance of Section c). However, our Wyner-Ziv coder outperforms H.26L FGS coding when there are packet losses in the base layer; and the higher the rate of the uncorrupted enhancement layer, the larger the performance gap between these two coders. The average PSNR performance gain of layered WZC increases with the PDU loss rate first, reaching 1.5 dB when the PDU loss rate is 6%. Because the amount of FEC is fixed (at 20%), the performance gain of layered WZC slightly decreases as the PDU loss rate (hence the residual RTP packet loss rate) goes further up to make the enhancement layer more corrupted.

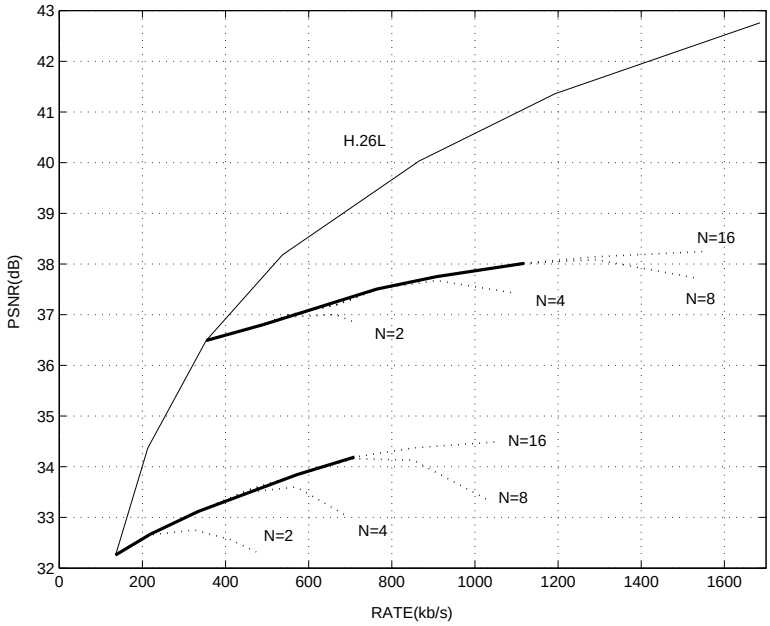
The decoded 10th frames of Football by H.26L FGS and layered Wyner-Ziv video coding (from the 7th simulated transmission) are shown in Fig. 12. It is easy to see that the decoded frame in Fig. 12 (b) has higher visual quality than that in Fig. 12 (a).

Similar simulations are also run on the CIF Foreman and Mother_daughter sequences, and results included in Fig. 11. For Foreman, the H.26L base layer is encoded at 305 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 255 Kb/s. The total transmission rate is 700 Kb/s. With 20% FEC, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are

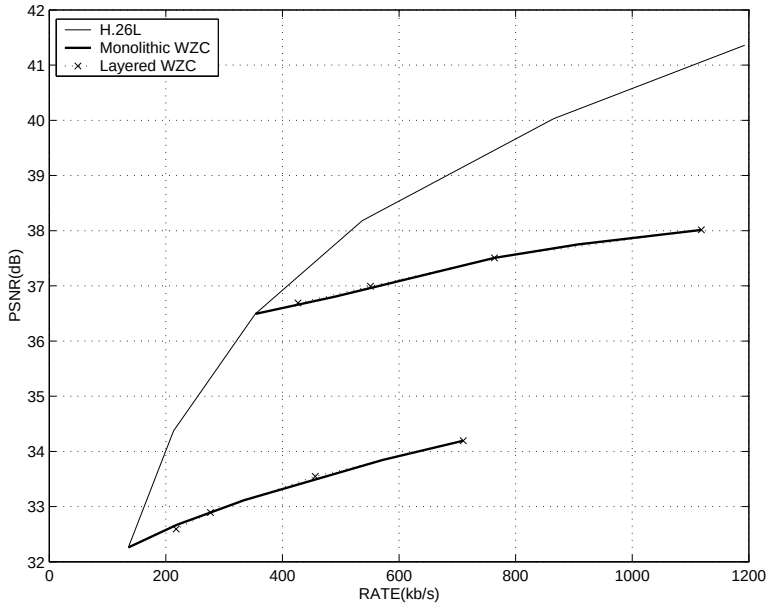
0.33%, 0.95%, 2.20%, 3.84%, and 5.66%, respectively. Again, 200 video transmissions are simulated for each PDU loss rate. The average PSNR performance gains of layered WZC over H.26L FGS are 0.67 dB, 0.9 dB, and 0.77 dB when the PDU loss rate is 6%, 8%, and 10%, respectively.

For Mother_daughter, the base layer is encoded at 257 Kb/s and the bit rate for the enhancement layer is 143 Kb/s. The total transmission rate is 500 Kb/s. With 20% FEC, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are 0.47%, 1.13%, 2.49%, 4.03%, and 5.68%, respectively. After 200 simulated video transmissions (at each RTP packet loss rate), the average PSNR performance gains of layered WZC over H.26L FGS are 0.25 dB, 0.29 dB, and 0.3 dB when the PDU loss rate is 6%, 8%, and 10%, respectively.

From both Figs. 10 and 11, we see that layered WZC is more error robust than H.26L FGS in video streaming applications. In addition, we see that more performance gain in terms of average PSNR is obtained for high motion sequences like Football. This is because the error-drifting problem becomes worse for high motion sequences in standard DPCM-based video coding; on the other hand, the distributed nature of WZC makes it a promising and viable technique for alleviating the effect of error drifting associated with standard coders.



(a)



(b)

Fig. 7. Illustration of successive refinement in our layered Wyner-Ziv video coder, assuming ideal SWC. (a) The operational rate-PSNR function of WZC is formed as the upper concave hull of different rate-PSNR points. (b) There is almost no performance loss between monolithic WZC and layered WZC.

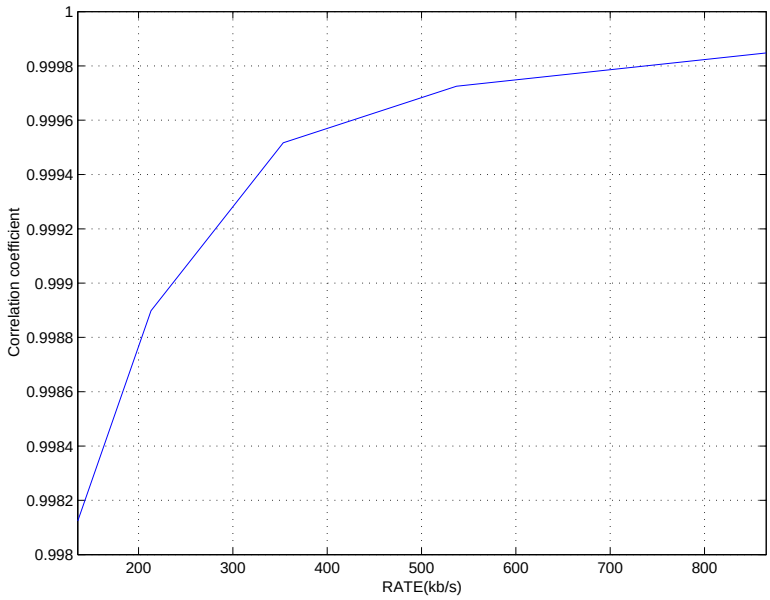


Fig. 8. The estimated correlation coefficient between the DC component of the original video and that of the side information for CIF Foreman. The horizontal axis represents the rate for the H.26L base layer.

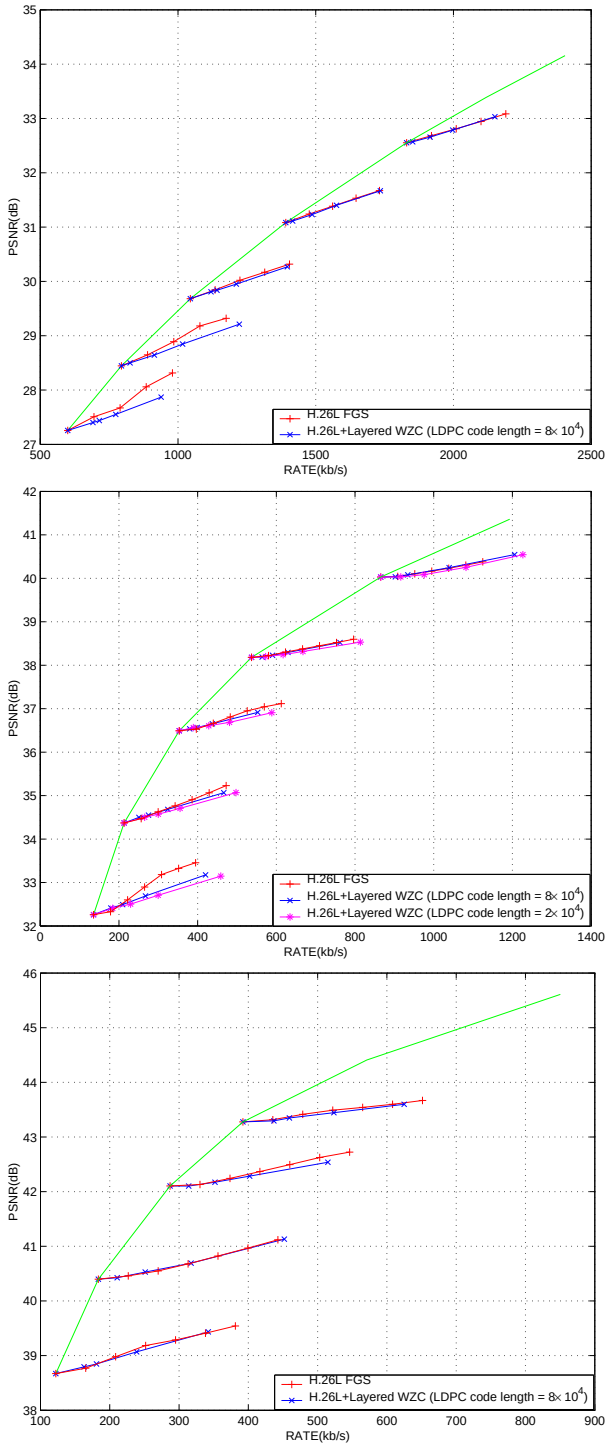


Fig. 9. Layered WZC of Football (top), CIF Foreman (middle), and Mother_daughter (bottom), starting from different "zero-rate" points. The sum of the rates for H.26L coding and WZC is shown in the horizontal axis.

Table I. The conditional entropy, LDPC code rate, and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane after NSQ of the DC (a) and the first two AC coefficients (b) and (c) after the DCT.

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.03433	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
1	0.03884	0.94	-	-
2	0.10157	0.85	$0.2137x + 0.2482x^2 + 0.0795x^3 + 0.0695x^7 + 0.3889x^8$	$0.5x^{24} + 0.5x^{25}$
3	0.19593	0.73	$0.2103x + 0.2062x^2 + 0.0615x^4 + 0.1667x^5 + 0.3002x^{13} + 0.0551x^{14}$	$0.5x^{14} + 0.5x^{15}$

(a)

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.00178	0.99	$0.2530x + 0.3067x^2 + 0.4403x^3$	$0.5x^{194} + 0.5x^{195}$
1	0.03768	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
2	0.02532	0.96	$0.1703x + 0.2714x^2 + 0.0136x^3 + 0.1046x^4 + 0.4400x^9$	$0.5x^{99} + 0.5x^{100}$
3	0.16754	0.77	$0.2313x + 0.2201x^2 + 0.1092x^3 + 0.3477x^8 + 0.0917x^9$	$0.5x^{15} + 0.5x^{16}$

(b)

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.02785	0.95	$0.1779x + 0.2605x^2 + 0.0843x^3 + 0.3452x^8 + 0.1322x^9$	$0.5x^{79} + 0.5x^{80}$
1	0.00488	0.98	$0.2695x + 0.3885x^2 + 0.3420x^3$	$0.5x^{134} + 0.5x^{135}$
2	0.03240	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
3	0.44291	0.49	$0.2330x + 0.1589x^2 + 0.0361x^3 + 0.0785x^5 + 0.1672x^6 + 0.0709x^{19} + 0.2554x^{20}$	$0.5x^7 + 0.5x^8$

(c)

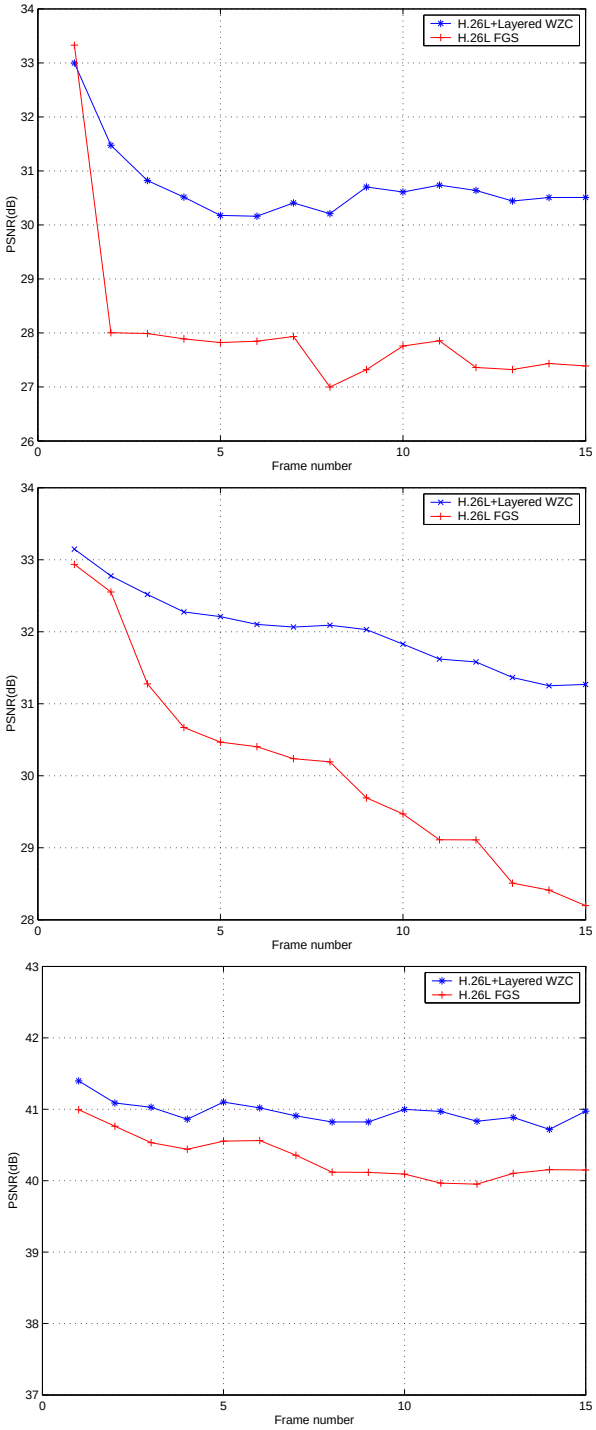


Fig. 10. Compared to FGS coding, Wyner-Ziv video coding offers substantial improvement in decoded video quality when the base layer (or decoder side information) suffers 1% macroblock loss for Football (top), 5% macroblock loss for Foreman (middle), and 5% macroblock loss for Mother_daughter (bottom).

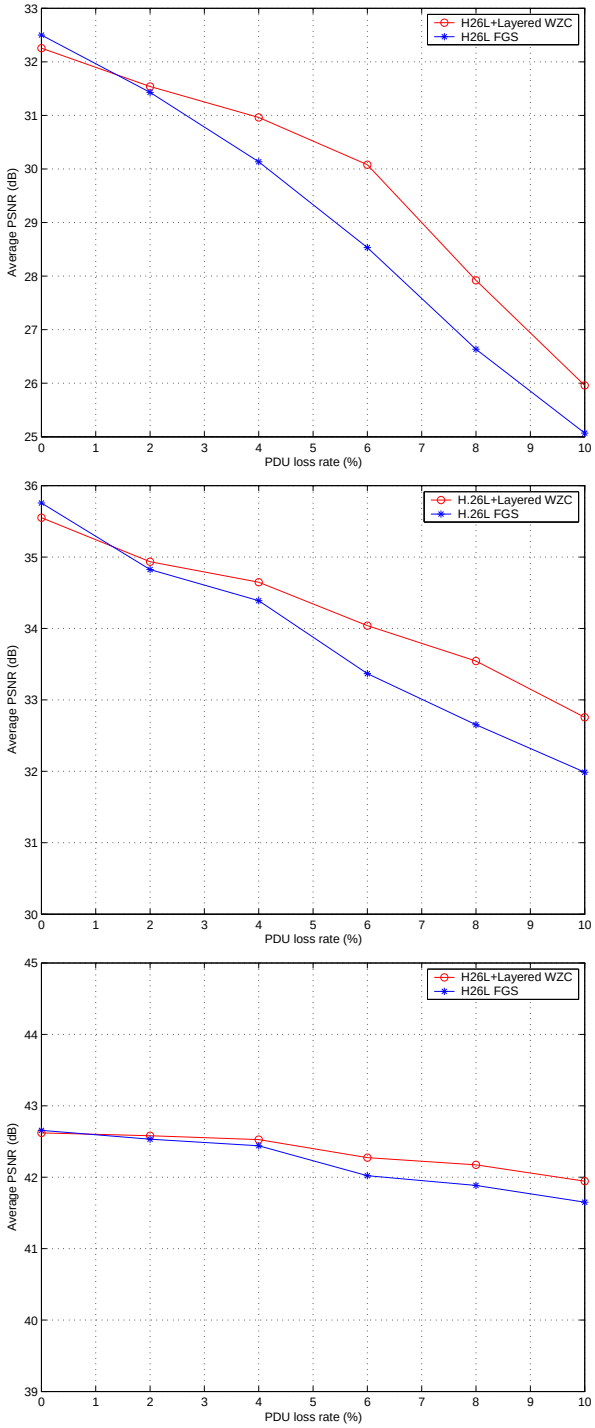
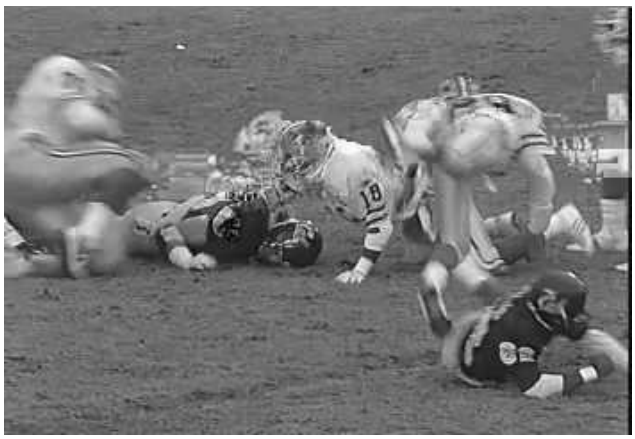


Fig. 11. Comparison of the Wyner-Ziv video coder and H.26L FGS coder when both are protected with RS-based FEC codes and transmitted over a simulated CDMA2000 1X channel for Football (top), CIF Foreman (middle) and Mother_daughter (bottom).



(a)

(b)

Fig. 12. Error robustness performance of Wyner-Ziv video coding compared with H.26L FGS for Football when both the base layer and enhancement layer bitstreams are protected with 20% RS-based FEC and transmitted over a simulated CDMA2000 1X channel with 6% PDU loss rate. The 10th decoded frame by (a) H.26L FGS and (b) Wyner-Ziv video coding in the 7th simulated transmission.

CHAPTER III

WYNER-ZIV VIDEO COMPRESSION AND FOUNTAIN CODES FOR RECEIVER-DRIVEN LAYERED MULTICAST *

A. Introduction

The increasing demand for video streaming over the Internet and third generation (3G) wireless networks has generated a lot of research interests in developing efficient and reliable multimedia delivery systems. In multicasting applications, live or pre-stored audio and video are simultaneously broadcast to potentially millions of clients over a network, where the communication channels between the sender and the clients are extremely diverse in available bandwidths and packet loss rates.

To efficiently address this heterogeneity, receiver-driven layered multicast (RLM) was proposed in [38] and further developed in [18]. In RLM, the encoded bitstream consists of a number of quality layers. Depending on the available bandwidth and packet loss rate, a client chooses the number of layers to *subscribe* to. Thus, RLM shifts rate control to the receiver side and avoids frequent asynchronous acknowledgements that cause packet collision and congestion. RLM is based on layered source and channel coding. For example, Chou *et al.* [18] combined embedded wavelet video coding and systematic RS codes. More recently, multiple description source coding based on RS coding [39], which offers effective packet loss protection, was combined with layered coding in [40] to multicast MPEG-4 FGS video over heterogeneous networks. However, there are two main problems in the schemes of [18, 40]. First, the

---

*©[2007] IEEE. Reprinted, with permission, from "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast" by Q. Xu, V. Stanković, and Z. Xiong, 2007. *IEEE Transactions on Circuits and Systems for Video Technology*, to appear.

standard MPEG-4 FGS coder [14] is very sensitive to packet loss in the base layer. Indeed, due to error drifting/propagation, a single packet loss can cause encoder-decoder mismatch and result in poor video reconstruction. Second, systematic RS codes, albeit being maximum distance separable, have high decoding complexity [19]. Therefore, they are not practical for time-constrained streaming and power-limited wireless applications.

Aiming at resolving the above problems, we propose a system for pre-stored video multicast over heterogeneous error-prone networks [41]. Instead of using MPEG-4 FGS [14], to reduce sensitivity to packet loss, we resort to Wyner-Ziv video coding [9]. Specifically, we employ the layered Wyner-Ziv video coder [20] discussed in the previous chapter, which forms a base layer using a standard video coder and treats it as decoder side information for generating an enhancement layer based on WZC. Since the enhancement layer is decodable with commensurate qualities at rates corresponding to layer boundaries, the coder of [20] resembles MPEG-4 FGS in terms of generating a base layer plus a scalable enhancement layer; however, the enhancement layer is formed "blindly" without the use of the base layer (the decoder side information), which alleviates problems such as error drifting/propagation associated with the DPCM-based MPEG-4 FGS and makes the Wyner-Ziv coder robust to packet loss in the base layer. However, the enhancement layer is a scalable bitstream sensitive to channel failures, which must be protected.

To reduce the decoding time and the computational complexity (the latter being crucial for power limited wireless devices), we choose digital fountain codes [17] over RS codes for error control [18]; the latter are maximum distance separable codes with order $n \log n$ encoding time and quadratic decoding time [19]. Fountain codes are sparse-graph codes that are ideally suited for multicast applications, because they are rateless in the sense of allowing a potentially limitless stream of output symbols

to be generated for a given input vector. In the RLM scenario, fountain codes form a "digital fountain" so that interested receivers can join the multicast group to "drink" from it just long enough to receive enough output symbols to recover the original video. Discovered by Luby, LT codes [17] are one of the first classes of efficient practical fountain codes for the erasure channel. Recently, the Raptor Type 10 (R10) FEC has been standardized into the 3GPP multimedia broadcast/multicast services (MBMS) and digital video broadcast-handheld (DVB-H) wireless networks.

In summary, to obtain error robustness, the system we propose combines the latest achievements in multimedia source and channel coding: error-resilient Wyner-Ziv video compression and rateless fountain codes. After Wyner-Ziv video coding each output enhancement layer is independently protected by a digital fountain code; each resulting bitstream is then packetized and sent to a separate multicast group. A receiver can subscribe to different multicast groups to dynamically adapt to variations of the bandwidth and packet loss rate. Instead of cyclically retransmitting the packets [42] generated with fixed rate erasure codes (e.g., RS codes), the sender keeps generating packets on the fly. Thus, the system enjoys the advantages of both WZC (robustness to packet loss in the base layer) and digital fountain coding (powerful erasure protection and rateless encoded bitstream). Our simulation results show a significant performance improvement over the scheme of [40].

The rest of the chapter is organized as follows. Section B gives a brief overview of the digital fountain codes. Section C explains our proposed scheme for RLM over the Internet and 3G wireless networks based on layered Wyner-Ziv video coding and digital fountain coding. The simulation results are presented in section D.

B. Fountain Codes

Fountain codes are rateless erasure codes that have near-capacity performance on packet erasure channels. For a given set of input symbols $(x_1, \ldots, x_k)$, a fountain code produces a potentially limitless stream of output symbols $z_1, z_2, \ldots$ Each output symbol is generated independently as the exclusive-OR sum of a randomly chosen set of input symbols. A decoding algorithm for a fountain code which can recover with high probability the original k input symbols from any set of n output symbols has an *overhead* of n/k. Fountain codes are called *universal* if they have fast encoding and decoding algorithms and overhead close to one for *any* erasure channel with erasure probability less than one.

The first universal fountain codes were developed by Luby, called the LT codes [17]. The encoding and decoding process is as follows. To generate an encoded symbol z_n, the encoder first randomly chooses the degree d_n from a degree distribution $\rho(d)$; then, it selects uniformly at random d_n distinct input symbols from $\{x_1, \ldots, x_k\}$ and sets z_n as their exclusive-OR sum. The encoding operation defines a Tanner graph connecting encoded symbols (check nodes) to source symbols (information nodes). The information on the degree of each received symbol and which source symbol it is connected to in the graph must be sent to the decoder. The decoder starts decoding by finding a check node z_j that is connected to only one information node x_i ($i \in \{1, \ldots, k\}$), sets $x_i = z_j$, and then adds x_i to all other check nodes that are connected to it; finally, all edges connected to x_i are removed. The above procedure is repeated until all information symbols are determined.

The degree distribution used for generating the output symbols lies at the heart

of LT codes. Luby proposed the ideal Soliton distribution [17] $\rho(1), \ldots, \rho(k)$, where

$$\rho(i) = \begin{cases} \frac{1}{k} & i = 1, \\ \frac{1}{i(i-1)} & i = 2, \ldots, k. \end{cases} \tag{3.1}$$

Although the ideal Soliton distribution is expected to work perfectly by heuristic analysis, it performs poorly in practice. However, it can be slightly modified to yield the robust Soliton distribution [17] in the following way. Let $R = c\ln(k/\delta)\sqrt{k}$ for some constant $c > 0$. Define

$$\tau(i) = \begin{cases} \frac{R}{ik} & i = 1, \ldots, \frac{k}{R} - 1, \\ \frac{R\ln(R/\delta)}{k} & i = \frac{k}{R}, \\ 0 & i = \frac{k}{R} + 1, \ldots, k, \end{cases} \tag{3.2}$$

then, for each i, add $\tau(i)$ to the ideal Soliton distribution $\rho(i)$ and normalize the sum to obtain the robust Soliton distribution: $\mu(i) = (\rho(i) + \tau(i))/\beta$, where $\beta = \Sigma_{i=1}^{k}[\rho(i) + \tau(i)]$. With the robust Soliton distribution, each encoded symbol can be generated independently on average by $O(\ln(k/\delta))$ symbol operations, and the k input symbols recovered from any $k+O(\sqrt{k}\ln^2(k/\delta)) = k(1+\varepsilon)$ encoded symbols with probability 1-δ after $O(k\ln(k/\delta))$ symbol operations on average. Thus, in contrast to the quadratic decoding complexity of RS codes, LT codes have almost linear decoding complexity[1]. The price paid for this complexity reduction is ϵ percent extra in rate.

C. Wyner-Ziv Video Coding and Fountain Codes for RLM

In this section we describe our system, which combines the Wyner-Ziv video coder [20] and LT fountain codes [17] for RLM. The block diagram of our proposed system

[1]The first class of fountain codes with linear decoding complexity are Raptor codes [43].

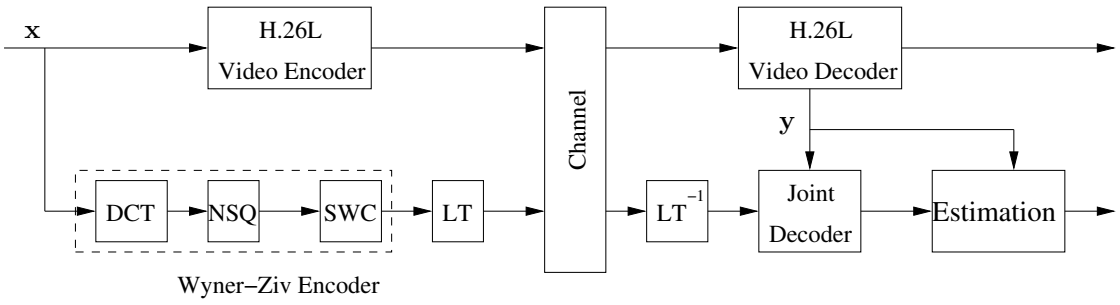


Fig. 13. Block diagram of the proposed system.

is given in Fig. 13. The system operates according to the following sequence of numbered steps:

1. The video encoder groups fixed number of frames into a GOF and generates for each GOF a base layer and L enhancement layers. Each enhancement layer is independently protected by an LT code [17], and the resulting bitstream is packetized into packets of length Q symbols each.

2. Prior to transmission of a GOF, a receiver, using its estimation of the bandwidth and packet loss rate, calculates the number of layers it can receive, i.e., the number of multicast groups it can subscribe to, in the way explained below. (We assume that a receiver has information about the statistics of the channel, obtained, for example, using already received packets. Note that such information is not needed at the transmitter.)

3. A sender encodes the j-th GOF into the base layer $b_{0,j}$ and L enhancement layers $e_{1,j}, \ldots, e_{L,j}$; each of these $L+1$ bitstreams is sent to a separate multicast group (see Fig. 14).

4. After receiving enough packets to reconstruct all the layers it subscribes to (for example, $i+1$ layers for GOF j, in Fig. 14), the receiver starts receiving packets

for the next GOF. However, if the receiver is not able to reconstruct any layer of the GOF, it stops receiving the packets for the current GOF and, while receiving data for the next GOF, performs error concealment (e.g., by simply repeating the last recovered frame of the previous GOF).

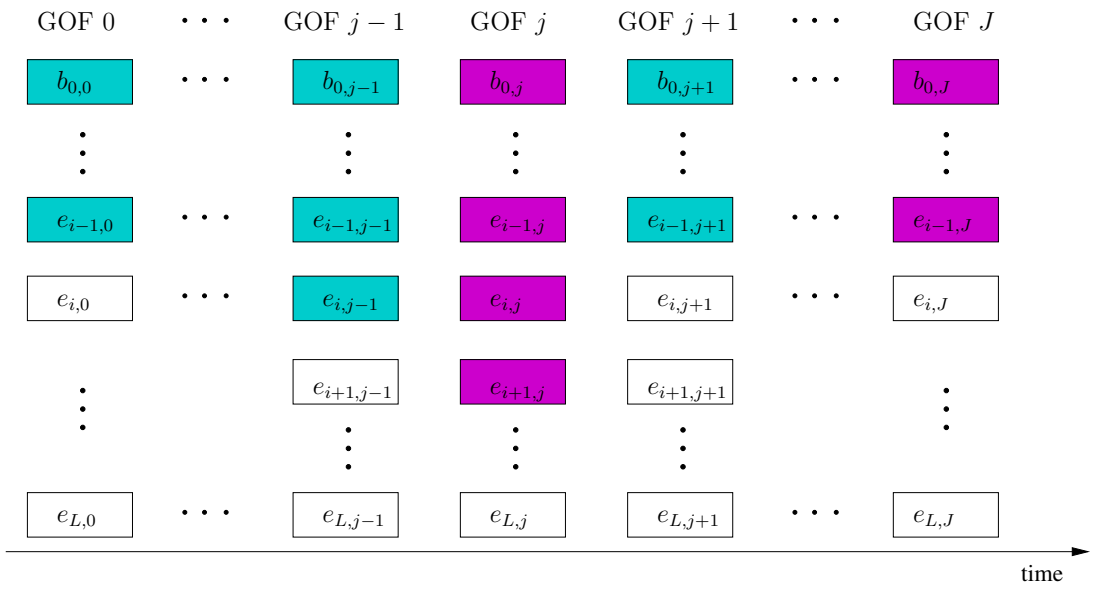


Fig. 14. Transmission of different GOFs.

Let K_q be the number of symbols in the q-th layer, $0 \leq q \leq L$ ($q = 0$ corresponds to the base layer). Then, approximately $K_q(1+\varepsilon)$ LT encoded symbols are needed for successful recovering of all K_q source symbols. Let T_d be the acceptable delay of a GOF; then, the maximum number of packets that the receiver can obtain for one GOF is BT_d, where B is the maximum transmission rate, or the available bandwidth (in packets per time unit). Let p be the average packet loss rate over a GOF; then, for each GOF, the receiver selects the multicast groups by determining maximum number of layers l, $0 \leq l \leq L$, that satisfies

$$\Sigma_{q=0}^{l} \frac{a_q}{1-p} \leq BT_d, \tag{3.3}$$

where $a_q = \lceil \frac{K_q(1+\varepsilon)}{Q} \rceil$. Note that the result $l < 0$ corresponds to the case when, for given B and p, the receiver cannot receive any layer.

For example, a pre-stored six-layer video sequence is to be transmitted to a large number of clients with different channel conditions. Suppose that $a_0 = a_1 = 80$ packets, $a_2 = a_3 = 100$ packets, and $a_4 = a_5 = 120$ packets. Assume that the acceptable delay T_d is 2 sec. Suppose that before the transmission of the first GOF, the receiver estimates the maximum available bandwidth to be $B = 170$ packets/sec and packet loss probability to be $p = 0.1$. Then using (3.3) the receiver will decide to subscribe to three multicast groups (to get the first three layers), i.e., $l = 2$ in (3.3). Then, the expected number of packets that will be received for this GOF is $BT_d(1 - p) = 170 \times 2 \times (1 - 0.1) = 306$, and the total number of packets needed for successful reconstruction is $\Sigma_{q=0}^{2} a_q = 260$. Suppose now that the receiver updates B and p to 200 packets/sec and 0.05, respectively; then using (3.3), it will decide to subscribe to four multicast groups for the next GOF. So, the average number of packets that can be received for the next GOF becomes $BT_d(1 - p) = 200 \times 2 \times (1 - 0.05) = 380$, and $\Sigma_{q=0}^{3} a_q = 360$ packets will be needed.

For real-time streaming applications, the acceptable delay T_d for a GOF should be no more than its duration T_{GOF}; thus, we impose the constraint $T_d \leq T_{GOF}$ to ensure continuous playback. The sender buffers live video frames as they arrive and then instantly encodes and packetizes them before transmitting to the clients, or directly transmit the pre-encoded packets in the case of multicasting pre-stored video. For both cases, continuous playback can begin after some initial start-up delay, which depends on the network bandwidth and decoding latency. In the case of *synchronous* multicast, all clients start to "listen" at the same time, and the sender successively transmits information about each GOF. In the scenario where the clients *asynchronously* start to "listen", the sender keeps sending each GOF at a separate

channel. (Finding the segmentation of a movie that minimizes starting delay is done in [44].) This transmission scheme allows the clients to randomly access any part of the video at any time, and video cassette recording functionalities, such as fast-forward/backward, can be realized at the receiver (with an initial delay of T_d) by switching to a different channel where a future or past GOF is transmitted. Instead of cyclically retransmitting the packets of a GOF generated with fixed-rate erasure codes (e.g., RS codes), the sender keeps generating encoded packets on the fly with the rateless LT codes. This avoids (with high probability) receiving the same LT encoded packets in two different transmissions of one GOF. So, if the receiver fails to reconstruct a GOF, it can stay at the current channel and collect the additional LT encoded packets of the same GOF without getting the packets that have already been received. If there is not enough number of channels for all the GOFs, the sender interleaves a number of successive GOFs and sends such a bitstream over one transmission channel. Note that this increases the initial start-up delay.

D. Experimental Results

In our simulations, we model the network as having no delay or latencies due to joining and leaving multicast groups. H.26L Test Model 9 video coderis used to generate the base layer (which is in the same time the side information at the Wyner-Ziv decoder for decoding the enhancement layers). 300 frames are compressed at a frame rate of 30 frames/sec for the standard CIF "Foreman" and "Mother_daughter" sequences. 20 frames are grouped and coded as one GOF each consisting of an I frame followed by 19 P frames. H.26L coder uses different quantization stepsizes to generate base layers at different compression ratios.

Enhancement layers are generated using WZC, where the correlation between

source X and side information Y is modeled as jointly Gaussian (in the DCT domain). After DCT of the original video, we only code the first three transform coefficients (i.e., DC and the first two AC coefficients), while the rest are discarded. For each coded transform coefficient, we use a four-bit nested scalar quantizer (with nesting ratio 16) to generate four bitplanes. The 12 bitplanes are then encoded using 12 different irregular LDPC codes (designed via density evolution), whose code rates are determined by the Slepian-Wolf limit [3]. This limit is computed as the conditional entropy rate of each coded bit plane of X given the side information Y for different GOFs (see [20] for details). For each GOF, the codeword lengths of the 12 LDPC codes are the same, but they vary from GOF to GOF (depending on the amount of motion), and range from 70 to 110 kilobits (kb). At the receiver, 100 iterations are used for LDPC iterative decoding to achieve the bit error rate of 5×10^{-5}. The base layer and $L = 4$ enhancement layers are formed for each GOF. The enhancement layers generated with the Wyner-Ziv encoder are independently protected with LT codes with robust Soliton distribution [17]; the resulting bitstreams are packetized into packets of length 200 bytes, and each is sent to a different multicast group. The receiver subscribes/unsubscribes to the multicast groups depending on the estimated maximum available bandwidth and the average packet loss rate.

First we give the rationale behind our experimental setup. Layered/FGS coding makes the basic assumption that the base layer is always correctly recovered at the decoder, so first we assume that the base layer is reconstructed perfectly at the decoder, and the enhancement layers are transmitted over a memoryless packet erasure channel for multicast. This will show the coding efficiency of layered Wyner-Ziv video coding and the erasure correction capability of the LT codes. Next, since we are targeting multicast applications, we compare our scheme to the best scheme of [40] which protects H.26L FGS video using multiple description scheme of [39] and

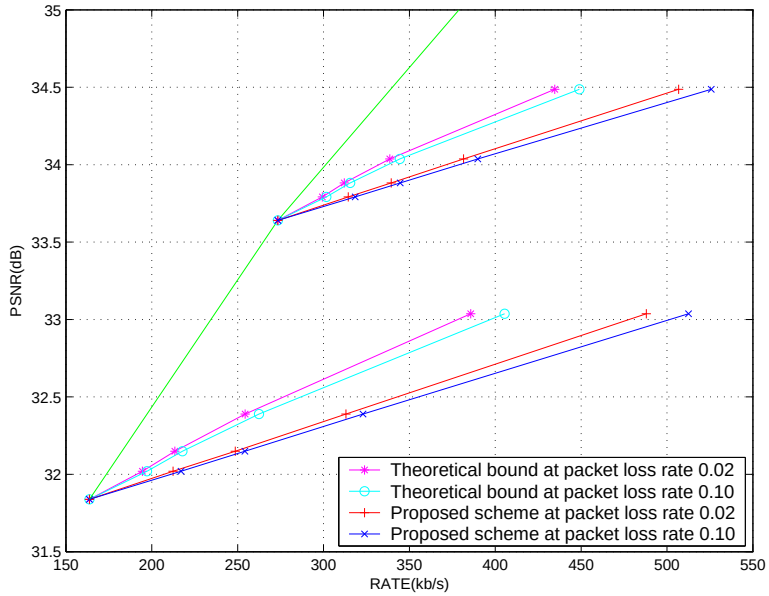
layered coding with RS codes.

Fig. 15 shows PSNR averaged over all 300 frames for the standard CIF "Foreman" and "Mother_daughter" sequences at two different packet loss rates for the enhancement layer and two bit rates of the base layer. The theoretical bounds with ideal SWC (in terms of achieving the Slepian-Wolf bound) and ideal (i.e., capacity-achieving) erasure protection coding over the same channels are shown as well. Compared to the theoretical limits, the loss due to our practical implementation with LDPC and LT codes is about 0.07 bit per sample for each frequency component. Thus the extra bit rate required for coding the first three frequency components is about $0.07 \times 3 \times 80000/\frac{2}{3} \sim 25$ kbps, assuming the codeword length of the LDPC code is 80,000 bits. Note that the overhead of the LT codes at code length 10,000 bits is approximately 1.07 in the experiments [45], that is, the receiver needs to receive about 7% more packets than with capacity-achieving maximum distance separable RS codes.

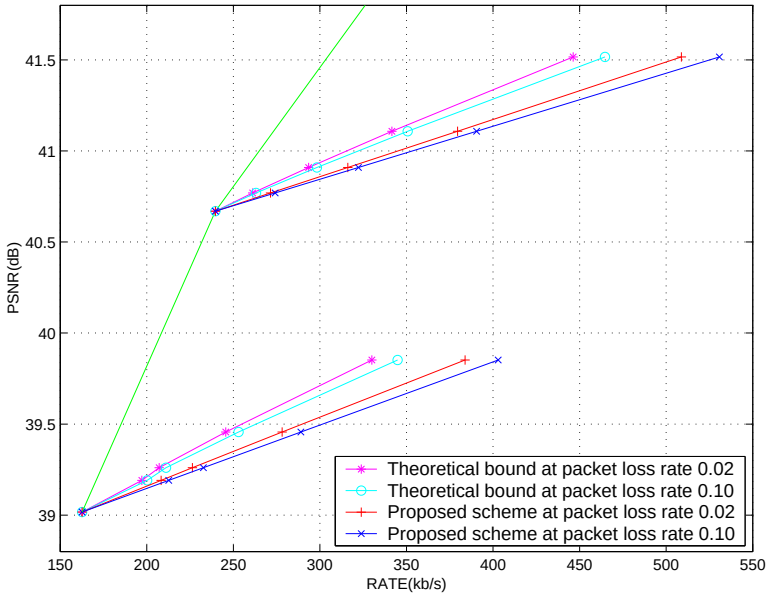
For comparison of multicasting schemes, we find the optimal source-channel symbol allocation as described in [40] assuming error-free base layer; the rates of the LDPC codes for SWC in our scheme are also chosen under the same assumption. We suppose that there are two sets of clients: low-bandwidth clients (LC) and high-bandwidth clients (HC). The unprotected base layer is sent to all clients over the same memoryless packet erasure channel. To simulate heterogeneous network, we assume that the enhancement layers are sent over different memoryless packet erasure channels with different available bandwidths, whose packet loss rates are 0.005 and 0.1 for LC and HC, respectively. This might be typical for a scenario where the clients with low packet loss rates are DSL subscribers and the clients with high packet loss rates are Internet users with wireless connection. This way, clients in different sets experience different channel conditions, and can therefore subscribe to different

number of multicast groups. Fig. 16 shows the obtained PSNR averaged over 100 independent transmissions of the first GOF vs. packet erasure rate of the base layer for the "Foreman" sequence. In each set of experiments, the available bandwidth for LC is always less than that for HC. For example, the dashed curves correspond to 341.6 kb/s for LC in Fig. 16 (a) and 351.2 kb/s for HC in Fig. 16 (b), while the dotted curves correspond to 372.8 kb/s for LC in Fig. 16 (a) and 389.6 kb/s for HC in Fig. 16 (b). If the packet loss rate in the base layer is zero, the coding performances for the two schemes are about the same. When there are packet losses in the base layer, besides being more suitable in practice (due to the employed rateless and low complex fountain codes), our scheme gives significantly better performance (up to 3 dB higher) than the one of [40], as seen from the figure.

This work represents the first step towards harnessing the inherent robustness property (to errors in the base layer) of our layered Wyner-Ziv video coder in video multicast by combining Wyner-Ziv video coding with LT codes. It has led to our more recent work on distributed joint source-channel coding of video [46], which will be presented in the next chapter.

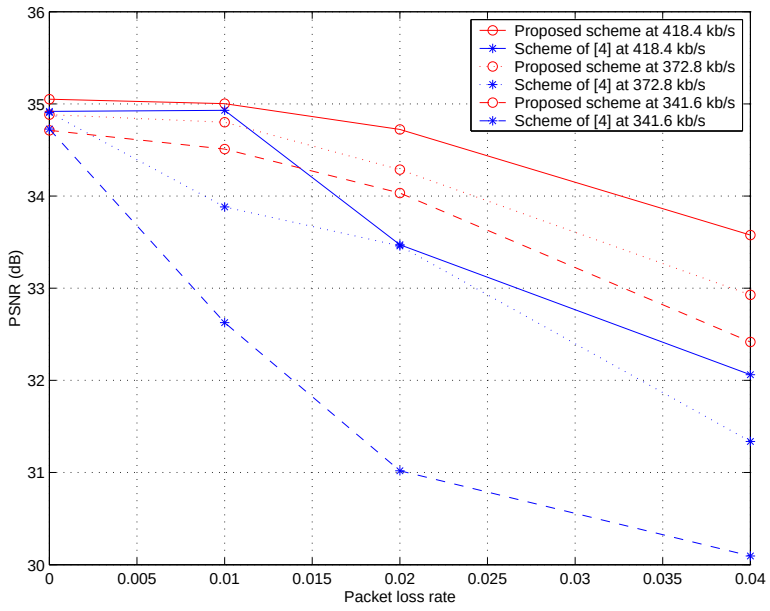


(a)

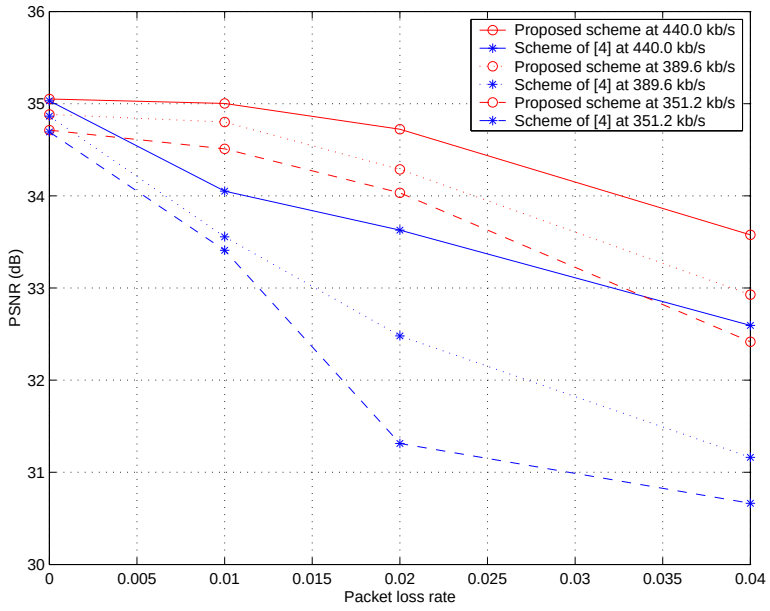


(b)

Fig. 15. Coding performance of the proposed scheme for: (a) the "Foreman" sequence, (b) the "Mother_daughter" sequence for two bit rates of the base layer and two packet loss rates.



(a)



(b)

Fig. 16. Average PSNR for 100 transmissions of the first GOF of the "Foreman" video sequence vs. packet loss rate of the base layer for: (a) LC (b) HC, and three different available bandwidths.

CHAPTER IV

DISTRIBUTED JOINT SOURCE-CHANNEL CODING OF VIDEO USING RAPTOR CODES *

A. Introduction

Multimedia communication over wireless networks has generated a lot of research interests in the past decade. Its main challenge lies in limited network bandwidth and the requirement of real-time playback on the one hand, and severe impairments of wireless links on the other. The additional issue has to do with the time-varying nature of wireless links and network heterogeneity, which make the channels between the sender and the clients extremely diverse in their available bandwidths and packet loss ratios. These diverse transmission conditions and bandwidth scarcity call for efficient *scalable* multimedia compression. Indeed, scalable video coding is expected to play a pivotal role in many emerging multimedia applications such as video broadcast/multicast over third generation (3G) wireless networks, interactive video, and wireless video surveillance networks. However, a scalable bitstream is usually very sensitive to channel noise as it suffers from error propagation. This is a limiting factor in their practical employment since wireless communication links are unreliable. Therefore, a *robust* scalable video coder is needed. Although standard video coders (e.g., H.264 [2]) can offer high coding efficiency in the scalable mode, they are very sensitive to packet loss, which results in error propagation/drifting.

Motivated by its potential applications in distributed sensor networks [6], video coding [8, 10, 9, 20], and compressing multi-spectral imagery [47], there has been a

---

*©[2007] IEEE. Reprinted, with permission, from "Distributed joint source-channel coding of video using raptor codes" by Q. Xu, V. Stanković, and Z. Xiong, 2007. *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 851-861.

flurry of research activities on DSC [6, 7] (e.g., SWC [3], WZC [4], and multiterminal source coding [48]) recently. For example, several efficient SWC and WZC schemes have been developed based on advanced channel coding for distributed compression (see [6, 7, 49, 50] and references therein). Moreover, Wyner-Ziv video coding [8, 10, 9, 20] has been proposed as a promising new technique. For example, a scalable video coder based on successive refinement for the Wyner-Ziv problem [12, 13] was presented in [20], where a standard decoded base layer was used as the decoder side information, and a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer.

The main advantage of Wyner-Ziv video coding over standard video coding (e.g., MPEG-4 FGS [14]) lies in error robustness. Specifically, the MPEG-4 FGS encoder generates the enhancement layer by coding the difference between the original video and the base layer reconstruction; then the decoder reconstructs the original video by adding an enhancement layer to the recovered base layer. This requires that the base layer recovered at the decoder is identical to that generated at the encoder. Thus, lost symbols in the base layer will cause the loss of synchronization between the encoder and decoder and result in severe error propagation. On the other hand, it is known from [4] that in WZC of quadratic Gaussian sources, separate encoding with joint decoding is as efficient as joint encoding (with the side information being present at both the encoder and decoder). Therefore, with Wyner-Ziv video coding, the enhancement layer can be generated "blindly" at the encoder without using the base layer (as side information). This way, transmission errors in the base layer will less likely cause encoder-decoder mismatch and hence have less impact on the reconstruction. This alleviates the problem of error drifting/propagation associated with FGS coding and makes the Wyner-Ziv video coder robust to errors/erasures in the base layer, as demonstrated in [20]. However, the *layered* enhancement bitstream

is very sensitive to transmission failures, since the channel is assumed to be noiseless in DSC in general and WZC in particular.

This chapter considers transporting Wyner-Ziv coded video over packet erasure channels and addresses distributed source-channel coding. Like in classic source-channel coding, although separation theorems [51, 11] have been shown to hold asymptotically (e.g., with infinite code length, delay and complexity), we show that distributed joint source-channel coding (JSCC)[1] outperforms separate source-channel coding in practice. Specifically,

1) We develop a design for distributed JSCC over packet erasure channels by extending the works on Slepian-Wolf coded nested lattice quantization [50] for WZC of quadratic Gaussian sources and on layered Wyner-Ziv video coding [20]. Instead of using separate channel codes for Slepian-Wolf compression (after quantization in WZC) and for protection, we adopt a *single* channel code for both SWC/compression and erasure protection in a distributed JSCC framework.

2) We make the specific choice of Raptor codes [43, 52] for the new application of distributed JSCC. Raptor codes are the latest addition to a family of low-complexity digital fountain codes [53], capable of achieving near-capacity erasure protection. They are proposed in [43] as precoded LT codes [17], and commonly used with LDPC precoding.

3) We employ a special class of LDPC codes called irregular repeat-accumulate (IRA) codes [54] as the precode for our Raptor code − a key factor in our successful code design. The IRA precoder is followed by an LT code which guarantees the rateless property of the overall Raptor code, meaning a limitless stream of packets can be generated by the encoder; thus the decoder can always receive enough packets

[1]Throughout this chapter, distributed JSCC means JSCC with decoder side information.

(in non-delay sensitive applications) for correct decoding, regardless of the packet loss ratio.

4) We state the design goal of our Raptor encoder, which is to minimize the number of packets the decoder has to receive[2] for correct decoding beyond the Slepian-Wolf compression limit. To this end, we vary the rate of the IRA precode and more importantly, introduce a bias towards selecting the IRA parity bits when making the random connections in forming the sparse-graph of the LT code. This bias is motivated by the fact that a correlated version of the IRA systematic bits is already available as side information at the decoder, and its optimization is embedded in the overall Raptor encoder design[3].

5) For the decoder design, due to the presence of decoder side information, we deviate from standard Raptor decoding and devise a new iterative soft-decision decoder that combines the received packets and the side information to perform joint decoding.

6) Our extensive simulations show that, compared to a separate design using WZC plus additional erasure protection, the proposed design provides better video quality.

The rest of the chapter is organized as follows. In Section B we give theoretical background on SWC, WZC, source-channel coding with decoder side information, and summarize two prior works that lead to the current one. In Section C, we review erasure protection coding techniques, ranging from RS codes to Tornado codes to digital fountain codes to Raptor codes. Section D describes two practical approaches

[2]We consider a packet as "received" only when it reaches the decoder without any error.

[3]We note that after our publication of [55], a similar idea was exploited in [56] in the context of unequal error protection with Raptor codes.

to SWC before focusing on distributed source-channel coding and pointing out advantages of a joint source-channel code design over a separate one. Section E describes our proposed video coder based on Raptor codes. Section F presents experimental comparisons between the proposed joint design, one separate design that uses WZC plus additional erasure protection, and another separate channel code design based on FGS source coding.

B. Theoretical Background and Related Works

First, a word about notation. Random variables are denoted by capital letters, e.g., X, Y. Realizations of random vectors of finite length n bits are denoted by bold-face lower-case letters, e.g., $\boldsymbol{x}$, $\boldsymbol{y}$. Matrices are denoted by bold-face upper-case letters; $\boldsymbol{I}_k$ and $\boldsymbol{O}_{k_1 \times k_2}$ are $k \times k$ identity matrix and $k_1 \times k_2$ all-zero matrix, respectively. All variables and channel codes are binary.

Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be a sequence of independent drawings of a pair of independent, identically distributed (i.i.d.) correlated random variables (X, Y). It is convenient to model the correlation between X and Y by a "virtual" correlation channel: $X = Y + N$, where the random variable N is the correlation channel noise that is independent of Y.

1. SWC

SWC is concerned with compressing discrete random variables X and Y separately and transmitting the resulting bitstreams over a noiseless channel to the receiver for joint decoding. The Slepian-Wolf theorem [3] asserts that if X and Y are compressed at rates R_X and R_Y, respectively, where $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$, and $R_X + R_Y \geq H(X, Y)$, then the joint decoder can recover them near losslessly. In the

sequel, we only focus on the special case, known as *source coding with decoder side information*, where Y is perfectly known at the decoder as side information. This case can be viewed as approaching the corner point $(R_X, R_Y) = (H(X|Y), H(Y))$ on the Slepian-Wolf rate region.

2. WZC

WZC [4] generalizes the setup of SWC in that coding of X is with respect to a fidelity criterion rather than lossless. In addition, the source X could be either discrete or continuous. The work of [4] examines the question of how many bits are needed to encode the source X under the constraint that the average distortion between X and decoded version $\hat{X}$ satisfies $E\{d(X, \hat{X})\} \leq D$, assuming that the side information Y (discrete or continuous) is available only at the decoder. Denote $R^*_{WZ}(D)$ as the achievable lower bound of the bit rate for an expected distortion D for WZC, and $R^*_{X|Y}(D)$ as the R-D function of coding X with side information Y available also at the encoder.

In general there is a rate loss associated with WZC, that is: $R^*_{WZ}(D) \geq R^*_{X|Y}(D)$. However, $R^*_{WZ}(D) = R^*_{X|Y}(D)$ when X and Y are zero-mean and jointly Gaussian and the distortion measure is MSE [4]. We restrict ourselves to this *quadratic Gaussian case* in WZC because there is no rate loss and it is of special interest in practice, where many image and video sources can be modeled as jointly Gaussian after mean subtraction.

3. Source-channel Coding with Decoder Side Information

When the transmission channel is noisy in the SWC problems, error protection is needed. In the noisy channel SWC case, the separation theorem is proved in [51] where it is shown that if the receiver has side information Y of the uncoded source

X, then the entropy of the source, $H(X)$, in the standard separation theorem is replaced by $H(X|Y)$. Equivalently, the Slepian-Wolf limit in this noisy channel case is $H(X|Y)/C$, where $C \leq 1$ is the channel capacity.

A separation theorem for lossy source-channel coding with decoder side information, i.e., the noisy channel WZC case, is given in [11]. It replaces the conditional entropy $H(X|Y)$ in the separation theorem for noisy channel SWC [51] by the Wyner-Ziv R-D function $R^*_{WZ}(D)$.

4. Related Works

A framework based on Slepian-Wolf coded quantization (SWCQ) is put forth in [50] for the quadratic Gaussian Wyner-Ziv problem. It is shown that the performance gap of high-resolution SWCQ to the Wyner-Ziv distortion-rate function $D^*_{WZ}(R)$ is exactly the same as that of high-rate classic source coding to the distortion-rate function $D(R)$. That is: with ideal SWC (or rate computed as $H(X|Y)$, where X is the *quantized* version of the Gaussian source), one-dimensional/two-dimensional SWCQ performs 1.53/1.36 dB away from $D^*_{WZ}(R)$ for quadratic Gaussian sources at high resolution.[4] Practical designs of one- and two-dimensional nested lattice quantizers together with multi-level LDPC codes for SWC and non-linear MSE estimation at the decoder give performance close to the theoretical limits of SWCQ.

Building upon the works of [13, 50] on WZC of ideal/Gaussian sources, a practical layered Wyner-Ziv video coder is proposed in [20] using the DCT, NSQ, and irregular LDPC codes for SWC. Denote the current frame of the original video as $\mathbf{x_c}$, which is

[4] Although the side information Y in Slepian and Wolf's original setup [3] has to be discrete, it can be discrete or continuous [57] in the special case of source coding with side information.

encoded with H.26L[5] to obtain the base layer (or decoder side information) $\mathbf{y}$. The DCT is used as an approximation to the cKLT [22], which makes the coefficients of the transformed block $\mathbf{X_c}$ of the original video $\mathbf{x_c}$ conditionally independent given the same transformed block $\mathbf{Y}$ of the side information $\mathbf{y}$. Each frequency component of $\mathbf{Y}$ (denoted by Y) acts as the decoder side information for the corresponding component of $\mathbf{X_c}$ (denoted by X_c). It is also assumed that X_c and Y are jointly Gaussian with $X_c = Y + N$, where N is zero-mean Gaussian and independent of Y, so that SWCQ is optimal for WZC (although DCT coefficients of images/video are better modeled as Laplacian distributed [34]). NSQ is a binning scheme that assigns the input DCT coefficients X_c to cosets and outputs only the coset indices X. Due to the correlation between X_c and Y, there still remains correlation between the coset index X and the side information Y. Multi-level LDPC codes are employed in SWC to compress the bit planes of X, ideally to the Slepian-Wolf rate limit of $H(X|Y)$.

C. Erasure Protection Coding

In this section, we review erasure protection codes, starting from the well-known RS codes and ending with Raptor codes – the latest in the family of digital fountain codes.

Systematic Reed-Solomon codes: Error protection over packet erasure channels can be realized with capacity-achieving RS codes. RS codes belong to a class of the so-called maximum-distance separable (MDS) codes, meaning that an $(n+r, n)$ RS code can recover the whole information sequence from any subset of n received symbols (provided that the erasure locations are known). However, the decoding complexity of practical $(n + r, n)$ RS codes is $O((n + r)^2)$ [58], making them too complex for

[5]H.26L refers to a dated implementation of the now well-known H.264 standard.

real-time applications.

Tornado codes: A new class of erasure protection codes, tornado codes, was introduced in [19]. By transmitting just below the channel capacity, hence sacrificing the MDS property, tornado codes can be encoded and decoded with linear complexity.

Digital fountain LT codes: Developed from tornado codes, digital fountain codes [53] are the latest in erasure correction coding. They are sparse-graph codes that are ideally suited for data protection against packet loss; they are rateless, in the sense of allowing a potentially limitless stream of output symbols to be generated for a given input sequence. A decoding algorithm for a fountain code, which can recover with high probability the original n input symbols from any set of $n + r$ output symbols, has the *overhead* of $\frac{r}{n} > 0$. Note that in MDS RS coding the overhead is always zero. A fountain code is called universal if it has fast encoding and decoding algorithms and the overhead close to zero for *any* erasure channel with erasure probability less than one. The first practical universal fountain code is the LT code [17]. LT coding is based on a Tanner graph connecting encoded symbols (check nodes) to source symbols (information nodes). The encoder generates an output symbol z_i by randomly choosing the degree d_i from a predetermined degree distribution and selecting uniformly at random d_i distinct source symbols from $x_1, \ldots, x_n$; z_i is then set as their XOR sum. The decoder first finds a check node z_j that is connected to only one information node x_i, $i \in \{1, \ldots, n\}$, sets $x_i = z_j$, adds x_i to all check nodes that are connected to it, and removes all edges connected to node x_i. This procedure is repeated until all information symbols are determined. For any $\delta > 0$, an LT code with the robust soliton distribution [17] can generate each encoded symbol independently on average by $O(\ln(\frac{n}{\delta}))$ symbol operations and recover the n input symbols from any $n+O(\sqrt{n}\ln^2(\frac{n}{\delta}))$ encoded symbols with probability of error δ after $O(n\ln(\frac{n}{\delta}))$ symbol operations on average.

Raptor codes: To decrease the encoding complexity, the average degree of the encoded symbols, which is $O(\ln n)$ for LT codes, should be reduced to a constant. Raptor codes [43] realize this goal by introducing a precoding step. Namely, to protect n input symbols, the decoding graph of an LT code must have the order of $n\ln(n)$ edges to ensure that all n input nodes are covered with high probability [17]; hence, one cannot encode at a constant cost if the number of collected output symbols is close to n. To circumvent this, a Raptor code first precodes the n input symbols with a fixed high-rate systematic linear code (e.g., LDPC code). Then the resulting precoded bitstream is fed to the LT encoder. Since now only a fraction of the precoded bitstream is needed for reconstructing the source, the $O(\ln n)$ bound on the average degree no long applies. With an appropriate design [43], for a given integer n and any real $\epsilon > 0$, a Raptor code can produce a potentially infinite stream of symbols such that any subset of symbols of size $n(1+\epsilon)$ is sufficient to recover the original n symbols with high probability. The degree of each encoded symbol is $O(-\ln\epsilon)$ and decoding time is $O(-n\ln\epsilon)$. Raptor codes currently give the best approximation of a digital fountain [53]. A potentially limitless sequence of packets can be generated on the fly after some small initial preprocessing with a linear encoding complexity. Decoding can be done in linear time after receiving just a few more than n encoding packets. Raptor codes are superior to the best LT codes not only over erasure channels, but also over the binary symmetric and additive white Gaussian noise channels [59].

D. Separate vs. Joint Design for Distributed Source-channel Coding

Since SWC is an integral part of the SWCQ framework for WZC, in this section, we first give an overview of practical SWC based on channel coding; we then provide extensions to the case when the Slepian-Wolf coded bitstream (after quantization in

WZC) has to be transmitted over a packet erasure channel − a scenario that calls for distributed source-channel coding. We present a code design where SWC and erasure protection are done separately and a joint design which performs SWC and erasure protection jointly.

1. Practical SWC

The proof of the Slepian-Wolf limit $H(X|Y)$ [3] is based on random binning, thus non-constructive. We review next two approaches proposed for practical SWC based on structured (or algebraic) binning [5].

Using the idea of Slepian and Wolf, Wyner [29] outlined a constructive binning scheme using channel codes for SWC, where each bin is a coset of a *good* parity-check code indexed by its syndrome. To compress the *binary* source X^n, a *syndrome-based encoder* employs a linear (n, k) channel code $\mathcal{C}$, given by its generator matrix $\boldsymbol{G}_{k\times n} = \begin{bmatrix}\boldsymbol{I}_k & \boldsymbol{P}_{k\times(n-k)}\end{bmatrix}$. (For simplicity we assume that $\mathcal{C}$ is systematic.) The corresponding $(n-k)\times n$ parity matrix is given by $\boldsymbol{H} = \begin{bmatrix}\boldsymbol{P}^T_{k\times(n-k)} & \boldsymbol{I}_{n-k}\end{bmatrix}$. Then, the encoder forms an $(n-k)$-length syndrome vector $\boldsymbol{s} = \boldsymbol{x}\boldsymbol{H}^T$ and sends it to the decoder. The decoder generates an n-length vector $\boldsymbol{t} = [\boldsymbol{O}_{1\times k} \quad \boldsymbol{s}]$ by appending k zeros to the received syndrome. Note that $\boldsymbol{c} = \boldsymbol{x} \oplus \boldsymbol{t}$ is a valid codeword of $\mathcal{C}$, where $\oplus$ denotes the XOR operator. By decoding $\boldsymbol{t} \oplus \boldsymbol{y}$ on $\mathcal{C}$, a codeword $\hat{\boldsymbol{c}}$ is obtained, and the source is reconstructed as $\hat{\boldsymbol{x}} = \hat{\boldsymbol{c}} \oplus \boldsymbol{t}$. To satisfy the Slepian-Wolf limit, we must ensure $\frac{n-k}{n} \geq H(X|Y)$. The syndrome-based approach [29] is optimal for SWC under the additive and independent noise models, since if the code $\mathcal{C}$ approaches the capacity of the "virtual" correlation channel $X = Y + N$, it also approaches the Slepian-Wolf limit.

In the above approach, each bin is indexed by a syndrome of a channel code. However, one can instead use parity-check bits to index the bins. We call this approach

parity-based binning. To compress source X^n, a *parity-based encoder* employs a linear $(n+r, n)$ systematic channel code $\mathcal{C}^p$ with generator matrix $\boldsymbol{G^p}_{n\times(n+r)} = [\boldsymbol{I}_n \quad \boldsymbol{P^p}_{n\times r}]$. The encoder forms an r-length parity vector as $\boldsymbol{p} = \boldsymbol{x}\boldsymbol{P^p}$ and transmits it to the decoder. The decoder generates an $(n+r)$-length vector $\boldsymbol{t^p} = \begin{bmatrix}\boldsymbol{y}_{1\times n} & \boldsymbol{p}\end{bmatrix}$, and by decoding $\boldsymbol{t^p}$ on $\mathcal{C}^p$, it obtains $\hat{\boldsymbol{c^p}} = \hat{\boldsymbol{x}}\boldsymbol{G^p}$, whose systematic part is the source reconstruction $\hat{\boldsymbol{x}}$. If the code $\mathcal{C}^p$ approaches the capacity of the "virtual" correlation channel, it also approaches the Slepian-Wolf limit. The Slepian-Wolf theorem mandates that $\frac{r}{n} \geq H(X|Y)$. To achieve the same amount of compression with both the syndrome- and parity-based approaches, the code rates of the employed codes $\mathcal{C}$ and $\mathcal{C}^p$ should be such that $r = n - k$. Then the two approaches are equivalent and generate the same encoder output if $\boldsymbol{H}^T = \boldsymbol{P^p}$. However, note that the parity-based approach has to employ a code with longer length, resulting in increased design complexity while not improving the compression efficiency. We thus conclude that for the SWC problem, in which the compressed bitstream is assumed to be perfectly available at the decoder, the syndrome-based approach is a better choice.

2. Transmission over Packet Erasure Channels

When the transmission channel for conveying the Slepian-Wolf compressed bitstream is noisy, source-channel coding with decoder side information is needed. This gives rise to the problem of distributed source-channel coding (see Section II-C).

The output of a syndrome-based Slepian-Wolf encoder are syndrome bits of a channel code, which are used here purely for compression, not for error protection. Therefore, when the transmission channel is noisy, following the separation principle [51, 11], first one channel code should be used to perform Slepian-Wolf compression and then the resulting syndrome bits protected by another channel code against errors introduced by the noisy transmission channel. The syndrome-based approach

for SWC can only be used in *separate* designs of the source and channel coding components. Such a separate design was proposed in [60] based on LDPC codes for SWC and digital fountain LT codes [17] for erasure protection. Although the separation approach is asymptotically optimal, joint designs are expected to perform better in practice.

Since SWC is essentially a channel coding problem [7], it is natural to combine the two channel codes − one for SWC and another for channel coding − into one single channel code for distributed JSCC. This can be achieved with the parity-based approach in SWC, because the graph structure of a channel code based on parity bits is more efficient than that based on syndrome bits in the presence of erasures. Indeed, if the amount of generated parity bits increases above the Slepian-Wolf limit, the extra redundancy can be exploited for protection. We thus view the source-channel coded bits as the parity bits of a systematic channel code and consider an equivalent channel coding problem over two parallel channels. The first channel is the noisy transmission channel through which the output bits of the encoder are transmitted, and it describes the distortion experienced by the parity bits of the code. The second channel is the "virtual" correlation channel between the source (the systematic bits of the channel code) and the decoder side information. This idea was previously exploited in [61, 62, 63, 64, 65] to design practical Slepian-Wolf codes for transmission over binary symmetric, Gaussian, and fading channels.

However, when the actual transmission channel is erasure based, designing a single channel code for joint SWC and erasure protection is difficult because a good candidate code should perform well over parallel concatenations of the correlation channel between X and Y and the packet erasure transmission channel. The search for such a good candidate code leads us to Raptor codes [43]. A precode of the Raptor code is a linear systematic $(n + r, n)$ code given by generator matrix $\boldsymbol{G^p}_{n\times(n+r)}$.

The encoder first forms an $(n+r)$-length codeword as $\boldsymbol{x}_s = \boldsymbol{x} \times \boldsymbol{G^p}$. Then, the output symbols are generated as $\boldsymbol{z} = \boldsymbol{x_s} \times \boldsymbol{S}^T$, where $\boldsymbol{S}$ is an $r' \times (n+r)$ matrix whose rows are sampled independently from the employed LT code degree distribution [17, 43]. Assuming that the capacity of the packet erasure channel is C, we must have $r' \geq nH(X|Y)/C$ [11], or more precisely, $r' \geq nH(X|Y)(1+\varepsilon)/C$, where ε is the Raptor code overhead. Note that there is not an upper bound on r', since the encoder can generate output symbols until the decoder receives enough to successfully decode; that is, the encoder can extend matrix $\boldsymbol{S}^T$ by generating new columns on the fly. The encoder output vector can be expressed as $\boldsymbol{z} = \boldsymbol{x} \times (\boldsymbol{G^p} \times \boldsymbol{S}^T)$, where the $n \times r'$ matrix $\boldsymbol{G^p} \times \boldsymbol{S}^T$ can be seen as a parity matrix of a bigger $(n+r', n)$ systematic code given by the generator matrix $\left[\boldsymbol{I}_n \ (\boldsymbol{G^p} \times \boldsymbol{S}^T)\right]$. Decoding starts when at least $nH(X|Y)(1+\varepsilon)$ bits are received and is done *jointly* on the whole decoding graph (see details in Section E).

We point out that a separate design based on concatenating a syndrome-based Slepian-Wolf code $\mathcal{C}$ with an LT code and a joint design with a Raptor code based on $\mathcal{C}^p$ precoding and an LT code are equivalent if: 1) the employed LT codes in both designs are the same; 2) $\boldsymbol{H}^T = \boldsymbol{P^p}$; 3) all LT parity bits of the Raptor code are connected to the parity bits of $\mathcal{C}^p$. Since the joint design based on Raptor codes does not have to be constrained by 3), there is obviously more freedom in the Raptor code construction, leading to improved performance over separate designs.

E. Distributed Joint Source-channel Coding of Video Using Raptor Codes

The block diagram of the proposed system for distributed JSCC of video is shown in Fig. 17. The video sequence is first encoded at a low bitrate with a standard video coder (H.26L [66] in our experiments) to generate a base layer, which is transmitted

to the receiver. At the receiver, the base layer is decoded and reconstructed; as in [20], we denote by Y the DCT coefficients of the reconstructed base layer, which will play the role of decoder side information. To improve the reconstruction video quality, the encoder then generates enhancement layers using WZC, or more precisely distributed JSCC. The rationale behind lies in correlation between the DCT coefficients X_c of the enhancement layer and the DCT coefficients of the reconstructed base layer Y, which we model with a "virtual" correlation channel, that is $X_c = Y + N$ [20], where N is an i.i.d. Gaussian random variable independent of Y (see Section 4). The DCT coefficients X_c of the original video sequence are first quantized to X, and then a multi-level Raptor code with IRA precoding is employed not only to compress further X (by exploiting remaining correlation between X and Y via the SWC binning scheme), but also to provide erasure protection (see Section 2).

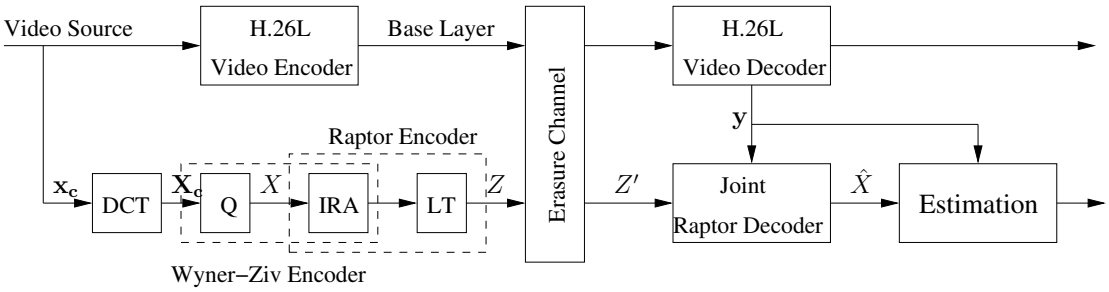


Fig. 17. Block diagram of our proposed video coder with Raptor codes. DCT denotes Discrete Cosine Transform, and Q stands for quantization.

We note that because Raptor codes have not been employed for JSCC with decoder side information before, there are several new issues with using them for distributed JSCC.

- First, in conventional erasure protection coding of n-length binary source sequence $X^n = \{x_1, \ldots, x_n\}$ with Raptor codes, a minimum of $n(1+\epsilon)$ output

symbols are needed for successful decoding (with high probability), where ϵ is the overhead percentage [43, 17]. However, *in distributed JSCC, the design goal of the encoder is to guarantee that only a minimum of* $nH(X|Y)(1+\varepsilon)$ *symbols are required at the decoder for successful decoding, where the* n*-length* $Y^n = \{y_1, \ldots, y_n\}$ *is the decoder side information.*

- Second, in contrast to hard-decision decoding[6] of conventional Raptor codes over erasure channels, the decoder side information necessitates iterative *soft-decision decoding*[7] in distributed JSCC to extract soft information from the Gaussian correlation channel.

The rest of this section describes how to resolve these issues by efficiently combining the received packets with side information Y in the proposed Raptor code design. Our key novelty lies in the choice of IRA (instead of conventional LDPC) precoding, which facilitates soft-decision decoding. For easy exposition, we assume that X is binary (with one-bit resolution in NSQ), although results reported in Section VI are based on using four-bit resolution in the quantizer, in which case joint decoding and estimation at the decoder are carried out similarly as described in detail in [20, 50].

1. Encoding

The proposed Raptor encoder with IRA precoding is depicted in Fig. 18 (a). First, the input binary sequence $X^n = \{x_1, \ldots, x_n\}$ is encoded with a systematic $(n+r, n)$ IRA precode, resulting in intermediate check symbols $u_1, \ldots, u_r$ and parity symbols

[6]By hard-decision decoding, we mean message-passing decoding [67] that passes "hard" information between iterations about whether a node is 0 or 1 and outputs hard decisions.

[7]By soft-decision decoding, we mean message-passing decoding [67] that passes "soft" information between iterations but outputs hard decisions.

$v_1, \ldots, v_r$. For $j = 1, \ldots, r$, u_j is the XOR sum of all input systematic symbols it is connected to, and v_j is computed as $v_j = v_{j-1} \oplus u_j$, with $v_0 = 0$ [54]. Then, the potentially limitless output stream $z_1, \ldots, z_m, \ldots$ is generated from the $(n+r)$-length sequence $x_1, \ldots, x_n, v_1, \ldots, v_r$ by encoding with an LT code.

Note that the $(n+r, n)$ IRA precode is not employed solely for SWC, but it also facilities protection against erasures. Therefore, it is not obvious that its code rate $\frac{n}{n+r}$ should be related to the Slepian-Wolf limit $H(X|Y)$ via $\frac{r}{n} \geq H(X|Y)$, leading to $\frac{n}{n+r} \leq \frac{1}{1+H(X|Y)}$ as in the separate design that employs the IRA code for SWC and an additional erasure protection code. The optimal IRA code rate now depends not only on the Slepian-Wolf limit (i.e., the correlation between X and Y), but also on the particular bipartite graph of the LT code.

Each LT output symbol z_i is connected randomly to d_i IRA systematic and parity symbols, where d_i is chosen from the LT code degree distribution [17]. In conventional Raptor encoding, systematic and parity symbols of the precode are treated equally in this random selection for LT coding. This means that each LT output symbol is connected with equal probability to any (systematic or parity) IRA symbol − thus all IRA symbols have in average the same degree (the number of connections to output LT symbols, i.e., the same number of parity-check equations involved in). Since the degree of an LT input symbol (IRA symbol) determines how strong it is protected against erasures, all IRA coded symbols in conventional Raptor coding are equally protected.

However, in our system, the decoder side information Y provides *a priori* knowledge about the IRA systematic symbols, and the decoder does not have such information about the IRA parity symbols. Consequently, if we apply conventional Raptor encoding with equal degrees of all IRA symbols, IRA parity symbols at the decoder would be almost useless since the systematic symbols would be easier recovered di-

rectly from the received LT packets due to side information. *In order to take full advantage of the IRA parity symbols, we introduce a bias towards selecting IRA parity symbols versus systematic symbols in forming the bipartite graph of the LT code.* This is done by selecting IRA parity symbols with probability $p > 0.5$ for a given LT output symbol. Note that in conventional Raptor encoding, $p = 0.5$. This way, we balance the effective realized protection between IRA systematic and parity symbols. The key challenge is to select the optimal p so that the improved protection of the parity symbols compensates presence of the side information for systematic symbols and thus maximizes performance.

The optimal p clearly depends on the IRA precode rate, and these two parameters must be considered jointly. In our encoder design, we select p and IRA precode rate experimentally. We start with an $(n + r, n)$ IRA code of rate less than $\frac{1}{1+H(X|Y)}$, ensuring $\frac{r}{n} > H(X|Y)$, then p is chosen in our simulations to minimize the overhead percentage ϵ, i.e., the number of bits $n\frac{r}{n}(1 + \epsilon) = r(1 + \epsilon)$ that the decoder has to receive for correct decoding and quick convergence of the overall Raptor code. Given the determined p, we adjust the rate of the IRA precode to further improve the performance. Our experiments show that the Raptor code performance is more sensitive to the choice of p than the IRA precode rate.

2. Soft-decision Decoding

A bipartite graph used for our joint Raptor decoder is shown in Fig. 18 (b). Let m be the number of received symbols and u_j the check sum of t systematic symbols $x_{j_1}, \ldots, x_{j_t}$; then from $v_j = v_{j-1} \oplus u_j$, it follows that $x_{j_1} \oplus \cdots \oplus x_{j_t} \oplus v_{j-1} \oplus v_j = 0$. In other words, the intermediate check symbols u_j's can be set to zero and viewed as check sums of the connected systematic symbols $x_{j_1}, \ldots, x_{j_t}$ and IRA parity symbols v_{j-1} and v_j. Therefore, we can think of $\tilde{X}^{n+r} = \{x_1, \ldots, x_n, v_1, \ldots, v_r\}$ as the

extended sequence of input symbols and $\tilde{Z}^{r+m} = \{u_1 = 0, \ldots, u_r = 0, z_1, \ldots, z_m\}$ as the extended sequence of received symbols. Then decoding of $X^n = \{x_1, \ldots, x_n\}$ is based on the iterative message-passing algorithm [67] on the created bipartite graph in Fig. 18 (b), where variable and check nodes are associated with $\tilde{X}^{n+r}$ and $\tilde{Z}^{r+m}$, respectively.

The log likelihood ratios (LLR's) for the systematic symbols are computed using the side information $Y^n = \{y_1, \ldots, y_n\}$ (assuming the "virtual" correlation channel between X_c and Y), and since we have no *a priori* knowledge of the IRA parity symbols $v_1, \ldots, v_r$, the LLR's corresponding to them are initially set to zero. In each decoding iteration, messages or LLRs are passed from a variable node $\tilde{x} \in \tilde{X}^{n+r}$ to a check node $\tilde{z} \in \tilde{Z}^{r+m}$ as follows

$$msg(\tilde{x} \rightarrow \tilde{z}) = \sum_{w \neq \tilde{z}} msg(w \rightarrow \tilde{x}) + msg_0(\tilde{x}),$$

where $msg_0(\tilde{x})$ is the initial LLR of the variable node $\tilde{x}$. Then, messages are passed from a check node $\tilde{z}$ back to a variable node $\tilde{x}$ as

$$\tanh \frac{msg(\tilde{z} \rightarrow \tilde{x})}{2} = \tanh \frac{msg_0(\tilde{z})}{2} \prod_{w \neq \tilde{x}} \tanh \frac{msg(w \rightarrow \tilde{x})}{2},$$

where $msg_0(\tilde{z})$ is the initial LLR of the check node $\tilde{z}$ (i.e., if $\tilde{z} = 0$, then $msg_0(\tilde{z}) = +\infty$; otherwise, $msg_0(\tilde{z}) = -\infty$).

At the end of the above soft-decision decoding process, X^n is decoded as $\hat{X}^n$, and the optimal estimate of X_c given $\hat{X}$ and Y at the decoder is computed as the conditional mean $\hat{X}_c = E(Xc|\hat{X}, Y)$ before $\hat{\mathbf{X}}_\mathbf{c}$ is converted to the pixel domain $\hat{\mathbf{x}}_\mathbf{c}$ via the inverse DCT.

F. Experimental Results

In this section we report our experimental results obtained with the standard CIF Foreman and SIF Football sequences. We encode 300 frames for Foreman and 100 frames for Football at 30 f/s. The base layer (or decoder side information) is generated with the H.26L video coder [66]. Every 20 frames are grouped and coded as one GOF that consists of one I frame followed by 19 P frames. Enhancement layers are generated using WZC, where the correlation between source X and side information Y is modeled as jointly Gaussian. We assume that $X_c = Y + N$ in the DCT domain, where the side information $Y \sim N(0, \sigma_Y^2)$ and the quantization noise $N \sim N(0, \sigma_N^2)$ due to H.26L coding are independent.

Similar to [20, Fig. 7], we estimate σ_N^2 for each DCT coefficient based on the MSE between X_c and Y within each GOF. We only code the first three DCT coefficients (i.e., the DC and the first two AC coefficients AC1 and AC2) and discard the remaining ones in each block[8]. Each coded coefficient X_c is quantized to four bits using NSQ [5], leading to $X = B_0B_1B_2B_3$ in its binary representation, and a total of $3 \times 4 = 12$ Raptor codes.

Each IRA precode is designed using density evolution with Gaussian approximation [54]; the input length n of the 12 IRA codes for each GOF is the same, but it differs from GOF to GOF (in the range of 70-110 Kb, depending on the amount of motion in each GOF). The distribution of the LT code we use is from [43]; although this distribution is optimal for the binary erasure channel, it provides good performance for the Gaussian channels as well [59]. Each LT check node is connected to the r IRA parity nodes with the bias probability p and to the n systematic nodes

[8]We note that more than three DCT coefficients may be required to be coded at high bit rates.

with probability $1-p$. The resulting output bitstreams are grouped into packets of 200 bytes each and sent over the packet erasure channel with packet loss ratio 0.1 (i.e., with capacity $C = 0.9$). At the receiver, 100 iterations are used in joint Raptor decoding. We assume error-free decoding if the probability of decoding error is less than 5×10^{-5}.

1. Coding Performance with Perfect Base Layer

In this subsection, we assume that the base layer is perfectly reconstructed at the receiver, and compare the proposed joint design based on Raptor codes (with IRA precoding and soft-decision decoding) to a separate design scheme, which concatenates IRA codes (for Slepian-Wolf compression) and conventional LT codes (for erasure protection).

In the separate design, the rate $\frac{n}{n+r}$ of the $(n+r, n)$ binary IRA code for the most significant bit B_0 of X is chosen such that the SWC rate $\frac{r}{n}$ is larger than the corresponding precomputed Slepian-Wolf limit $H(B_0|Y)$, which depends on the joint statistics between B_0 and Y; and the IRA code rate for B_i $(i = 1, 2, 3)$ is picked so that the corresponding SWC rate $\frac{r}{n}$ is lower bounded by the Slepian-Wolf limit $H(B_i|B_0, \ldots, B_{i-1}, Y)$, which is computed from the joint statistics between B_i and $\{B_0, \ldots, B_{i-1}, Y\}$ [20].

In our joint design, the best probability p and the rate of the IRA precode for each of the 12 Raptor codes (for each GOF) are determined as described at the end of Section 1.

Table II lists our computed Slepian-Wolf limits, the actual SWC rates (as determined by $\frac{r}{n}$) we use in the separate design, and the corresponding $\frac{r}{n}$'s of the IRA precodes in the joint design, for the first GOF of Foreman (with the H.26L base layer rate being 273.41 Kb/s).

Fig. 19 shows the obtained PSNR (averaged over all 300 frames for Foreman and 100 frames for Football) as a function of the total transmission rate. For each sequence, two different H.26L base layer rates are chosen, and the enhancement layers are transmitted over the packet erasure channel (with $C = 0.9$). The four PSNR-rate points on each curve correspond to results after consecutively decoding packets generated for each of the four bit planes. The theoretical limit is $nH(X|Y)(1+\epsilon)/C$, where $H(X|Y) = H(B_0|Y) + H(B_1|B_0, Y) + H(B_2|B_0, B_1, Y) + H(B_3|B_0, B_1, B_2, Y)$ and ϵ is the overhead of the LT code in erasure correction coding. In the separate design, the minimum number of LT symbols that the encoder needs to transmit is defined as $nR_X(1+\epsilon)/C$, where R_X is the actual SWC rate (third column of Table II) achieved by IRA codes and the overhead of the LT code ϵ is around 7% in our experiments. The horizontal performance gap (in rate) between the separate design and the theoretical limit represents the loss due to practical SWC with IRA codes. For the DC component of the first GOF of Foreman, the IRA code profiles are $\lambda(x) = 0.7248x^2 + 0.2301x^6 + 0.0451x^7$, $\rho(x) = x^{25}$; $\lambda(x) = x^2$, $\rho(x) = x^{37}$; $\lambda(x) = 0.5936x^2 + 0.4064x^{12}$, $\rho(x) = x^{23}$; and $\lambda(x) = 0.4293x^2 + 0.2644x^9 + 0.3064x^{10}$, $\rho(x) = x^{12}$ for B_0, B_1, B_2, and B_3, respectively.

For our joint design based on Raptor codes, we report the best results optimized at $p = 0.8$ for the bias probability. It is seen from Fig. 19 that to achieve the same average PSNR, the number of transmitted packets with the joint Raptor code design is 7-9% and 5-6% less than that with the separate design for the CIF Foreman and SIF Football, respectively. For the DC component of the first GOF of Foreman, the IRA precode profiles are $\lambda(x) = 0.8088x^2 + 0.1912x^4$, $\rho(x) = x^{24}$; $\lambda(x) = x^2$, $\rho(x) = x^{42}$; $\lambda(x) = 0.6185x^2 + 0.1198x^{11} + 0.2617x^{12}$, $\rho(x) = x^{24}$; and $\lambda(x) = 0.3204x^2 + 0.1911x^9 + 0.2435x^{13} + 0.2449x^{22}$, $\rho(x) = x^{18}$ for B_0, B_1, B_2, and B_3, respectively. If we denote the minimum number of LT symbols the decoder has to receive for correct decoding

as $nR'_X(1+\epsilon')$, where the rate R'_X is the corresponding $\frac{r}{n}$ (fourth column of Table 1) of the IRA precode, based on results in the figure, the overhead ϵ' due to Raptor coding is 5-6%.

As theoretically predicted, our proposed joint design improves the performance of the separate design *by taking the advantages of the Raptor codes over the LT codes* in conventional erasure protection coding. Indeed, in LT coding, each parity-check symbol is connected randomly to a predetermined number of information symbols. In separate SWC and LT coding of X^n, it is possible that an information symbol is not connected to any of the received $n(1+\epsilon)$ LT parity-check symbols. It cannot then be recovered, although the probability of this event decreases as n increases. On the other hand, in our proposed joint design (as in Raptor coding), the additional connections to the information symbols (realized via precoding) reduce this probability.

2. Coding Performance with Corrupted Base Layer

In this subsection, we investigate robustness to the reconstruction errors in the base layer. The base layer and enhancement layers are generated by encoding the first 20 frames (one GOF) of the CIF Foreman and SIF Football sequences and transmitted over the same packet erasure channel. To illustrate improved robustness of Wyner-Ziv video coding over classic FGS video, besides the two designs (joint and separate) described in the previous subsection, we included another separate scheme based on H.26L FGS [15] video coding and erasure protection. Besides LT coding for erasure protection, we also consider MDS RS coding (RS codes are used purely to simplify implementation, there is no conceptual difference between RS codes and LT codes for erasure protection). To study the impact of MDS coding, the enhancement layers in the two separate designs are protected with either LT codes or with MDS RS codes. Thus, five different schemes are tested: 1) the proposed joint design based on Raptor

codes (with IRA precoding and LT codes), 2) the separate IRA + LT design, 3) the separate IRA + RS design, 4) H.26L FGS + LT, and 5) H.26L FGS + RS. Note that schemes 2 and 3 exploit IRA codes for Slepian-Wolf compression in WZC.

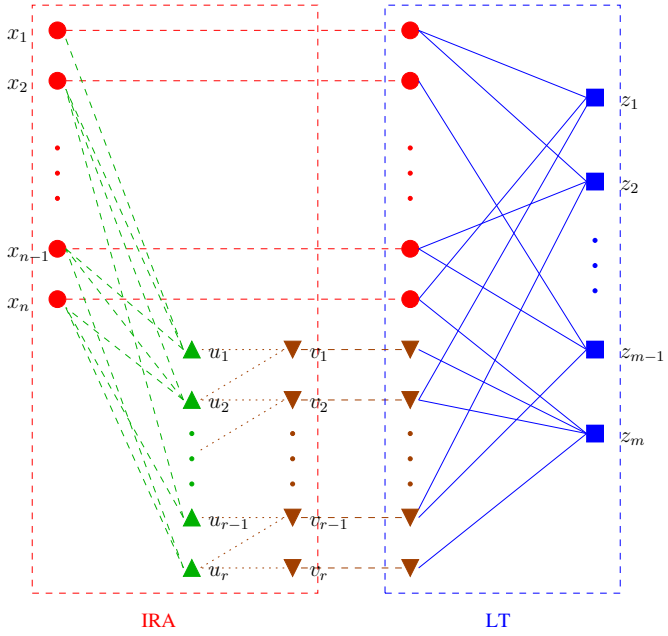
The base layer is encoded at 334.2 Kb/s and 1762 Kb/s, for the CIF Foreman and SIF Football sequence, respectively, where 10% of the rate is used for RS parity symbols; that is, an RS code of the rate $\frac{9}{10}$ is employed for erasure protection of the base layer. The bitrate of the enhancement layer is fixed at 281.1 Kb/s and 492.7 Kb/s, for the CIF Foreman and SIF Football sequence, respectively. The generated video packets are transmitted over a packet erasure channel where packet losses are introduced randomly with probability q.

In all experiments the LT code rate in schemes 2 and 4 is chosen to be 0.82 so that the probability of the LT decoding success is high at packet loss ratio 0.1. The code rates of the IRA code and LT code in scheme 1 are kept the same as in schemes 2 and 4 for fair comparisons, and the bias probability p in scheme 1 is set to be 0.8 as in the previous subsection. When RS codes are used, we employed the multiple description-based PET system of [39], which provides the most effective protection (at the expense of increased complexity and delay). The optimal source-channel symbol allocation (the RS code rates) is determined by using the fast unequal error protection algorithm of [68], assuming a packet loss ratio of 0.1.

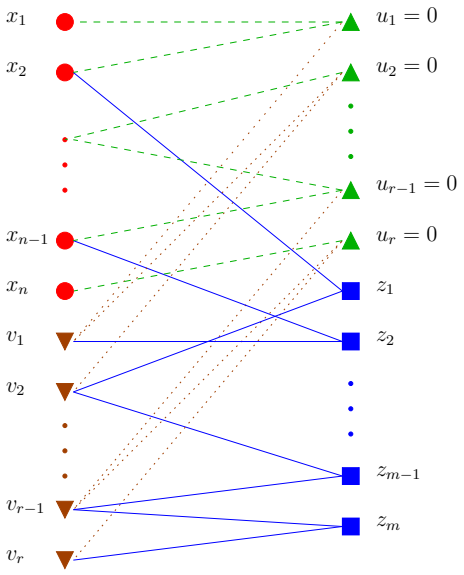
To evaluate robustness to the optimization mismatch (e.g., due to unknown channel statistics at the encoder), all five schemes are designed assuming channel packet loss ratio 0.1 and tested at five different loss ratios $q = 0.06$, 0.07, 0.08, 0.09, and 0.1. The obtained average PSNR over all 20 frames and one hundred simulations as a function of q is shown in Fig. 20. Note that after decoding, there are still residual errors in the base layer. For example, for the CIF Foreman sequence, residual packet loss ratios in the base layer were 0.47%, 1.03%, 2.13%, 3.41%, and 5.00%,

at $q = 0.06$, 0.07, 0.08, 0.09, and 0.1, respectively. (A simple error concealment is done during decoding.) For the enhancement layers, in schemes 2 and 3, the whole layer where the first syndrome decoding failure occurs is discarded together with all successive layers. This is done because Slepian-Wolf decoding cannot be performed with corrupted syndromes; therefore we must ensure that the entire bitstream fed to the Slepian-Wolf decoder is error free.

From the figure we can see that the joint scheme performs uniformly better than all separate design schemes (up to 1.2 dB and 1 dB, for the CIF Foreman and SIF Football, respectively, when compared to the separate scheme in its worst). The second conclusion from the figure is that the distributed coding schemes (schemes 1, 2, and 3) are more robust than FGS schemes in general, showing that employed WZC is capable of alleviating the effect of error drifting associated with standard FGS coding. We can also observe that the schemes with LT codes give a better reconstruction quality than the corresponding schemes based on RS codes at low packet loss ratios; we explain this by the fact that the schemes 3 and 5 are overprotected, since they are optimized for $q = 0.1$; that is, the LT code rates in schemes 2 and 4 are higher than the RS code rates in schemes 3 and 5. On the other hand, the LT-based schemes provide slightly worse quality at high packet loss ratios (where the optimization is performed), due to MDS property of RS codes.



(a)

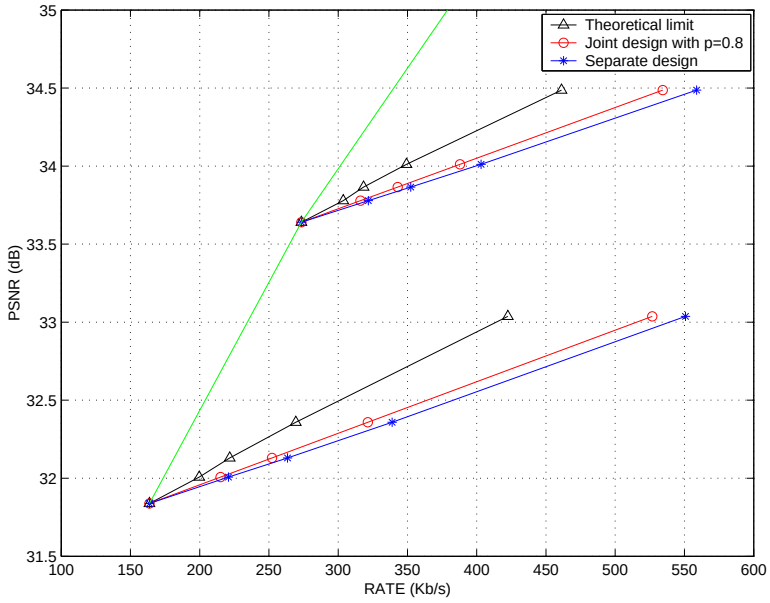


(b)

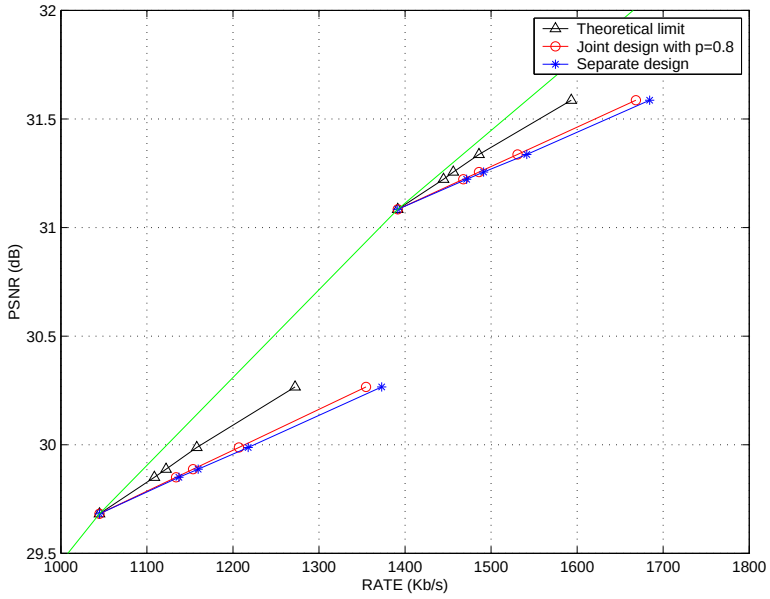
Fig. 18. (a) The graphical representation of the proposed Raptor encoder with IRA precoding. (b) The bipartite graph of our joint Raptor decoder.

Table II. Our computed Slepian-Wolf limits, the actual SWC rates (given by $\frac{r}{n}$) we use in the separate design, and the corresponding $\frac{r}{n}$'s of the IRA precodes in the joint design, for the DC, AC1 and AC2 of the first GOF of Foreman. The unit measure for all entries is bit.

	Slepian-Wolf limit	Actual SWC rate $\frac{r}{n}$ (separate design)	Corresponding $\frac{r}{n}$ (joint design)
DC: B_0	0.0942	0.14	0.12
B_1	0.0458	0.08	0.07
B_2	0.1173	0.17	0.16
B_3	0.2717	0.38	0.34
AC1: B_0	0.0837	0.12	0.11
B_1	0.0013	0.02	0.02
B_2	0.0131	0.03	0.03
B_3	0.1070	0.16	0.14
AC2: B_0	0.0027	0.02	0.02
B_1	0.0547	0.09	0.08
B_2	0.0329	0.06	0.06
B_3	0.1901	0.27	0.25

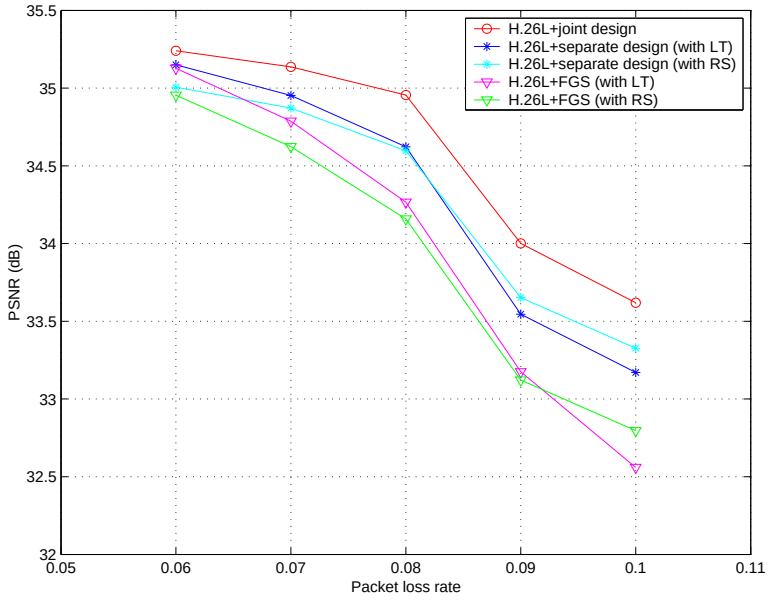


(a)

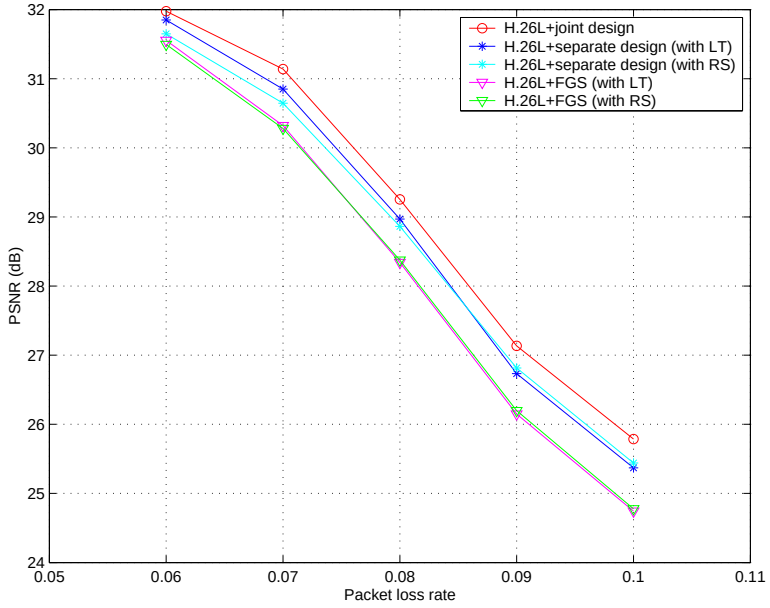


(b)

Fig. 19. Average PSNR (in dB) performance vs. bit rate (in Kb/s) between our distributed JSCC design and separate IRA and LT design for (a) the CIF Foreman and (b) the SIF Football sequences. The base layer is generated using H.26L and the packet loss ratio of the erasure channel is 0.1. The theoretical limited is $nH(X|Y)(1+\epsilon)/C$, with n being the input code length, $H(X|Y)$ the computed Slepian-Wolf limit, $\epsilon = 0.07$, and $C = 0.9$.



(a)



(b)

Fig. 20. Performance comparisons of the joint Raptor code design, separate IRA + LT design, separate IRA + RS design, H.26L FGS + LT, and H.26L + RS for (a) the CIF Foreman and (b) the SIF Football, as a function of the packet erasure rate. All schemes are designed for packet loss ratio 0.1.

CHAPTER V

CONCLUSIONS

The focus of the this dissertation is layered Wyner-Ziv video coding as a new approach to video compression and delivery. In Chapter II, we have proposed a practical layered Wyner-Ziv video coding system using the DCT, NSQ, and irregular LDPC code based SWC. The low-complexity DCT is used as an approximation to the cKLT. NSQ is the simplest nested quantization scheme that corresponds to quantization in classic source coding. LDPC code based SWC exploits the correlation between the quantized version of the source and the side information, and can be viewed as the counterpart of entropy coding in classic source coding. Our layered video coding system achieves scalability as the layered Wyner-Ziv bitstream enhances the standard base layer bitstream in such a way that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Simulation results demonstrate that layered WZC is more error robust than H.26L FGS coding in video streaming applications.

Although our results are encouraging, much remains to be done to make Wyner-Ziv video coding more viable in practice. First, more accurate statistical modeling will improve the quantizer and the SWC design. Second, LDPC code design can be further improved by using density evolution without the Gaussian approximation for different source distributions. In addition, it is desirable to improve the performance of LDPC codes with shorter block length to reduce the time delay in video coding. Our current implementation of layered WZC only allows decoding at layer boundaries – decoding at the middle of a layer (bit plane) will suffer a high performance loss as unavailable bits in the layer have to be treated as erasures. The counterpart of scalable source codes (e.g., arithmetic codes) or progressively decodable channel codes are needed for

scalable SWC to achieve fine-grained scalability. But scalable SWC remains to be a challenging problem.

We have relied on extensive simulations when testing error robustness of our layered Wyner-Ziv video coder, which is only designed under the assumption of *noiseless* channels. When the channel for the Wyner-Ziv enhancement bitstream is noisy, we have a problem of source channel coding, which is addressed in the following chapters. On the other hand, if the channel for the base layer bitstream is not perfect, the side information might be absent at the Wyner-Ziv decoder; although theoretical results for optimal encoding in this case appeared more than 20 years ago [69], we have not seen any corresponding practical code design even for ideal sources yet.

In Chapter III, we have proposed a system for real-time streaming of live or pre-stored video over heterogeneous error-prone networks based on layered Wyner-Ziv video coding and digital fountain coding. The former is employed to improve the error robustness to packet loss, while LT code is chosen over RS code for error control due to its rateless property and low decoding time although RS code is optimal in terms of recovering erasures.

For real-time video streaming which requires bounded decoding delay, transmission packets must be decoded with very low decoding latency so that the video can be played out continuously. In contrast to LT decoding, which has $O(k \ln(k/\delta))$ complexity, implementing iterative LDPC decoding with codeword length of 10^5 bits and 50 decoding iterations is time consuming (it takes about seven seconds of CPU time in our experiments on a PC having Pentium IV 2400 MHz processor). Reducing the LDPC codeword length decreases decoding time, but worsens the performance. Thus, this is currently a limiting factor in applying our system to live video streaming.

In Chapter IV, we extended the work in Chapter III to channel coding for distributed JSCC and expanded the powerful concept of digital fountain codes for erasure

protection in the process of accommodating the decoder side information. We have also developed a practical distributed JSCC scheme that exploits a single digital fountain Raptor code for both compression and protection for transmission over erasure channels. With this solution, we are able to reflect the advantages of Raptor codes over LT codes to the distributed coding case. Thus, the joint design based on our novel distributed JSCC paradigm is superior to designs where compression and protection coding are treated separately. In addition, while the separate design scheme has to wait until enough number of LT encoded symbols are collected for decoding of all Slepian-Wolf coded syndromes, in our proposed scheme, the decoding error gradually decreases as more encoded symbols become available. Our future work is to further optimize the IRA code as the precode using the extrinsic information transfer (EXIT) chart [70] to improve the coding performance of the proposed joint scheme.

REFERENCES

[1] T. Sikora, "MPEG digital video-coding standards," *IEEE Signal Process. Magazine*, vol. 14, pp. 82–100, Sep. 1997.

[2] T. Wiegand, G. Sullivan, G. Bjintegaard, and A. Luthra, "Overview of the h.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 13, pp. 560–576, Jul. 2003.

[3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, Jul. 1973.

[4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[5] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250–1276, Jun. 2002.

[6] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Magazine*, vol. 19, pp. 51–60, Mar. 2002.

[7] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Magazine*, vol. 21, pp. 80–94, Sep. 2004.

[8] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," *submitted to IEEE Trans. Image Process.*, 2006.

[9] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 93, pp. 71–83, Jan. 2005.

[10] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-Ziv coding of video: applications to error resilience," *IEEE Trans. Multimedia*, vol. 6, pp. 249–258, Apr. 2004.

[11] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[12] Y. Sterinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1636–1654, Aug. 2004.

[13] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," *IEEE Trans. Signal Process.*, vol. 53, pp. 3269–3281, Aug. 2005.

[14] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 11, pp. 301–317, Mar. 2001.

[15] Y. He, F. Wu, S. Li, Y. Zhong, and S. Zhang, "H.26L-based fine granularity scalable video coding," in *ISCAS*, Scottsdale, AZ, May. 2002, pp. 548–551.

[16] H. Garudadri, H. Chung, N. Srinivasamurthy, and P. Sagetong, "Video transport over wireless networks," in *Proc. 12th annual ACM international conference on Multimedia*, New York, NY, Jun. 2004, pp. 416–419.

[17] M. Luby, "LT codes," in *Proc. 43rd IEEE Symp. on the Foundations of Computer Science*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.

[18] P. Chou, A. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Trans. Multimedia*, vol. 3, pp. 108–122, Mar. 2001.

[19] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D.A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569–584, Feb. 2001.

[20] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," *IEEE Trans. Image Process.*, vol. 15, pp. 3791–3803, Dec. 2006.

[21] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440–442, Oct. 2002.

[22] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loeve transforms," in *Proc. DCC'03*, Snowbird, UT, Mar. 2003, pp. 283–292.

[23] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 849–866, Nov. 1998.

[24] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," in *Proc. VCIP'04: Special Session on Multimedia Technologies for Embedded Systems*, San Jose, CA, Jan. 2004, vol. 5308, pp. 83–91.

[25] A. Sehgal, A. Jagmohan, and N. Ahuja, "Scalable Wyner-Ziv coding using wyner-ziv codes," in *Proc. PCS'04*, San Francisco, CA, Dec. 2004, CD–ROM.

[26] H. Wang and A. Ortega, "WZS: Wyner-Ziv scalable predictive video coding," in *Proc. PCS'04*, San Francisco, CA, Dec. 2004, CD–ROM.

[27] M. Tagliasacchi, A. Majumdar, and K. Ramchandran, "A distributed-source-coding based robust spatio-temporal scalable video codec," in *Proc. PCS'04*, San Francisco, CA, Dec. 2004, CD–ROM.

[28] S.S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181–1203, May 2003.

[29] A. Wyner, "Recent results in the shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 2–10, Jan. 1974.

[30] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, Mar. 1999.

[31] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and gaussian-markov sources," *IEEE Trans. Communications*, vol. 38, pp. 82–93, Jan. 1990.

[32] J. M. Shapiro, "Embedded iamge coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, pp. 3445–3463, Dec. 1993.

[33] B.-J. Kim, Z. Xiong, and W. Pearlman, "Very low bit-rate embedded video coding with 3-d set partitioning in hierarchical trees (3-d SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 1365–1374, Dec. 2000.

[34] D. Reininger and J. Gibson, "Distribution of the two-dimensional DCT coefficients," *IEEE Trans. Communications*, vol. 31, pp. 835–839, Jun. 1983.

[35] T. Richardson, M. A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[36] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.

[37] Y. Wang and Q. Zhu, "Error control and concealment for video communications: A review," *Proc. of the IEEE*, vol. 86, pp. 974–997, May 1998.

[38] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in *Proc. ACM SIGCOMM 96*, Stanford, CA, Aug. 1996, pp. 117–130.

[39] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737–1744, Nov. 1996.

[40] P.A. Chou, H. Wang, and V. Padmanabhan, "Layered multiple description coding," in *Proc. 13th Int'l Packet Video Workshop*, Nantes, France, Apr. 2003, CD–ROM.

[41] Q. Xu, V. Stanković, and Z. Xiong, "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast," *IEEE Trans. Circuits Syst. Video Technol.*, to appear 2007.

[42] S. Floyd, V. Jacobson, C. Liu, S. McCanne, and L. Zhang, "A reliable multicast framework for light-weight sessions and application level framing," *IEEE/ACM Trans. Networking*, vol. 5, pp. 784–803, Dec. 1997.

[43] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, pp. 2551–2567, Jun. 2006.

[44] C. Huang, R. Janakiraman, and L. Xu, "Optimal loss-resilient media streaming using priority encoding," in *Proc. 12th Annual ACM Intern. Conf. Multimedia*, New York, NY, Oct. 2004, pp. 152–159.

[45] D. MacKay, *Information Theory, Inference & Learning Algorithms*, Chapter 50, Cambridge University Press, Cambridge, U.K., 2003.

[46] Q. Xu, V. Stanković, and Z. Xiong, "Distributed joint source-channel coding of video using raptor codes," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 851–861, May 2007.

[47] C. Tang, N. Cheung, A. Ortega, and C. Raghavendra, "Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: a distributed source coding approach," in *Proc. DCC'05*, Snowbird, UT, Mar 2005, pp. 437–446.

[48] T. Berger, *The Information Theory Approach to Communications*, Springer-Verlag, New York, 1977.

[49] V. Stanković, A.D. Liveris, Z. Xiong, and C.N. Georghiades, "On code design for the Wlepian-Wolf problem and lossless multiterminal networks," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1495–1507, Apr. 2006.

[50] Z. Liu, S. Cheng, A. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, pp. 4358–4379, Oct. 2006.

[51] S. Shamai and S. Verdu, "Capacity of channels with side information," *European Trans. Telecommunications*, vol. 6, pp. 587–600, Sept.-Oct. 1995.

[52] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor codes for reliable download delivery in wireless broadcast systems," *Consumer Communications and Networking Conference*, vol. 1, pp. 192–197, Jan. 2006.

[53] M. Mitzenmacher, "Digital fountains: a survey and look forward," in *Proc. ITW'04 IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004, pp. 271–276.

[54] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd Intl. Symp. Turbo codes and related topics*, Sept. 2000, pp. 1–8.

[55] Q. Xu, V. Stanković, and Z. Xiong, "Distributed joint source-channel coding for video using Raptor codes," in *Proc. DCC'05*, Snowbird, UT, Mar. 2005, pp. 491.

[56] N. Rahnavard and F. Fekri, "Finite-length unequal error protection rateless codes: design and analysis," in *Proc. IEEE Globecom'05,*, Saint Louis, MO, Nov. 2005, pp. 1353–1357.

[57] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Process.*, vol. 86, pp. 3123–3130, Nov. 2006.

[58] J. Bloemer, M. Kalfane, M. Karpinski, R. Karp, M. Luby, and D. Zuckerman, "An xor-based erasure-resilient coding scheme," Int. Computer Science Institute Technical Report, TR-95-48, 1995.

[59] R. Palanki and J. S. Yedidia, "Rateless codes on noisy channels," in *IEEE International Symposium on Inform. Theory*, Chicago, IL, Jun. 2004, pp. 37.

[60] Q. Xu, V. Stanković, and Z. Xiong, "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast," in *Proc. PCS'04*, San Francisco,

CA, Dec. 2004, CD–ROM.

[61] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. DCC'02*. Mar. 2002, pp. 252–261, Snowbird, UT.

[62] J. Garcia-Frias, "Joint source-channel decoding of correlated sources over noisy channels," in *Proc. DCC'01*, Snowbird, UT, Mar. 2001, pp. 283–292.

[63] P. Mitran and J. Bajcsy, "Turbo source coding: a noise-robust approach to data compression," in *Proc. DCC'02*, Snowbird, UT, Mar. 2002, pp. 465.

[64] A.D. Liveris, Z. Xiong, and C.N. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using ira codes," in *Proc. MMSP'02 IEEE Workshop on Multimedia Signal Process.*, St. Thomas, US Virgin Islands, Dec. 2002.

[65] M. Sartipi and F. Fekri, "Source and channel coding in wireless sensor networks using LDPC codes," in *Proc. First Annual IEEE Communications Society Conf. on Sensor Communications and Networks*, Santa Clara, CA, Oct. 2004, pp. 309–316.

[66] T. Wiegand, "H.26L test model long-term number 9 (tml-9) draft0," *ITU_T Video Coding Experts Group, VCEG-N83d1*, Dec. 2001.

[67] F. Kschischang, B. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[68] V. Stanković, R. Hamzaoui, and Z. Xiong, "Real-time error protection algorithms of embedded codes for packet erasure and fading channels," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 14, pp. 1064–1072, Aug. 2004.

[69] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 31, pp. 727–734, Nov. 1985.

[70] Y. Sun, A. Liveris, V. Stankovic, and Z. Xiong, "Near-capacity dirty-paper code designs based on TCQ and IRA codes," in *Proc. ISIT'05*, Adelaide, Australia, Sep. 2005, pp. 184–188.

VITA

Qian Xu was born in Jingzhou, China. She received the B.S. degree in Computer Science from the University of Science&Technology of China in 2002, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Texas A&M University in 2004 and 2007, respectively.

Upon admittance to Texas A&M, Ms. Xu was a recipient of the TxTEC Fellowship in September 2002. From 2001 to 2002, she worked as a student intern in the Internet Media group at Microsoft Research Asia, Beijing, China. She spent a summer in the Computational Biology Division at Translational Genomic Research Institute in 2006.

Her research interests include distributed video coding, video compression, pattern recognition, and genomic signal processing.

The typist for this dissertation was Qian Xu.